

Interrater Reliability of Osteopathic Sacral Palpatory Diagnostic Tests Among Osteopathy Students

Giacomo Consorti, DO (Italy); Fabio Basile, DO (Italy); Leonardo Pugliese, DO (Italy);
Marco Petracca, DO (Italy)

From the Centre pour l'Étude, la Recherche et la Diffusion Ostéopathiques in Rome, Italy (Consorti, Basile, Bugliese, and Petracca), and the San Pietro Fatebenefratelli Hospital in Rome, Italy (Petracca).

Financial Disclosures:
None reported.

Support: None reported.

Address correspondence to
Giacomo Consorti, DO (Italy),
Via Capitinzano, 33, 00178
Roma, Italy.

Email:
giacomo.consorti@gmail.com

Submitted
January 31, 2017;
revision received
November 24, 2017;
accepted
December 7, 2017.

Context: Somatic dysfunctions are a key element of osteopathic practice. The evaluation of somatic dysfunctions is achieved by assessment of the TART (tissue texture abnormality, asymmetry, restriction of motion, tenderness) parameters. The reliability of a diagnostic method is the crux of successful treatment. Interrater reliability of osteopathic palpatory diagnostic tests have been studied on different anatomical areas, but there are no studies on the evaluation of all of the TART parameters on the sacrum.

Objective: To evaluate the interrater reliability of osteopathic sacral palpatory diagnostic tests. The hypothesis was that 3 trained osteopathy students at the end of their curriculum could achieve at least moderate agreement on osteopathic sacral palpatory diagnostic tests.

Methods: Three students from the Centre pour l'Étude, la Recherche et la Diffusion Ostéopathiques school in Rome, Italy, at the end of their curriculum participated as raters and received consensus training. Eligible subjects among students of the same school were recruited on a voluntary basis to be tested. All of the raters tested the sacrum by evaluating the TART parameters on every subject for 3 minutes. Raters were blinded to the other raters' findings. Interrater reliability was evaluated using Fleiss κ statistics.

Results: Fifty-two subjects (20 women) were enrolled in the study. Mean (SD) age was 25.9 (7.03) years; height, 1.73 (0.09) m; weight, 68.73 (14.2) kg; and body mass index, 22.66 (3.58). Agreement was fair for tissue texture abnormality ($\kappa=0.28$), asymmetry ($\kappa=0.29$), restriction of motion ($\kappa=0.32$), and tenderness ($\kappa=0.34$); agreement was slight for landmark position ($\kappa=0.06$) and diagnosis of somatic dysfunction ($\kappa=0.17$).

Conclusion: Results showed a level of agreement ranging from slight to fair in the assessment of the TART parameters among raters, who were in their last year of osteopathy school. The tenderness parameter was the most reliable. Our findings are consistent with other interrater reliability studies carried out in different body regions, contributing to show an overall heterogeneous level of diagnostic reliability in osteopathy.

J Am Osteopath Assoc. 2018;118(10):637-644. Published online August 20, 2018.
doi:10.7556/jaoa.2018.132

Keywords: interexaminer reliability, interrater reliability, osteopathic sacral tests, palpatory diagnostic test, somatic dysfunction

Somatic dysfunctions are a key element of osteopathic practice. Somatic dysfunction is an impaired or altered function of related components of the somatic (body framework) system: skeletal, arthrodial, and myofascial structures, as well as related vascular, lymphatic, and neural elements.¹ The palpation diagnosis of somatic dysfunction and the use of osteopathic manipulation by osteopathic practitioners to relieve or ameliorate patient discomfort and pain are hallmarks of osteopathic principles and practice.²

TART is a mnemonic for the 4 diagnostic criteria of somatic dysfunction, as defined by the *Glossary of Osteopathic Terminology*¹:

- **Tissue texture abnormality:** A palpable change in tissues from skin to periarticular structures that represents any combination of the following signs: vasodilatation, edema, flaccidity, hypertonicity, contracture, fibrosis, as well as the following symptoms: itching, pain, tenderness, paresthesias. Types of [tissue texture abnormalities] include: bogginess, thickening, stringiness, ropiness, firmness (hardening), increased/decreased temperature and increased/decreased moisture.
- **Asymmetry:** Absence of symmetry of position or motion; dissimilarity in corresponding parts or organs on opposite sides of the body that are normally alike; of particular use when describing position or motion alteration resulting from somatic dysfunction.
- **Restriction of motion:** A resistance or impediment to movement.
- **Tenderness:** 1. Discomfort or pain elicited by the osteopathic practitioner through palpation. 2. A state of unusual sensitivity to touch or pressure.

The reliability of a diagnostic method is the crux of successful treatment. A diagnostic test should be reproducible in the same patient by 2 or more independent raters (or examiners), or at the very least, by the same rater on 2 separate occasions. These are referred to as

interrater (or interexaminer) and intrarater (or intraexaminer) reliability, respectively. For a finding to be clinically relevant, 2 independent raters should be able to agree on its presence or absence. If a diagnostic test does not satisfy this basic requirement, then it is considered unreliable.³ In the past decade, interrater reliability gained attention in osteopathic literature³⁻⁸ because of the perceived need for a stronger methodologic ground for osteopathic clinical research and practice.

One of the first attempts to assess interrater reliability in osteopathic medicine was performed by Upledger⁹ in 1977. His study showed substantial agreement in examining craniosacral motion, despite the fact that the examination was performed on children and used a poor methodology. In 2001, Moran and Gibbons¹⁰ showed that interrater reliability for simultaneous palpation at the head and the sacrum ranged from poor to nonexistent, with intraclass correlation coefficients ranging from -0.09 to 0.31. Degenhardt et al¹¹ studied the vertebral region and found differences between interrater reliability before and after consensus training. In the pretraining evaluation of interrater reliability, κ ranged from 0.02 to 0.34, which is within the poor to fair reliability range. After consensus training, reliability improved, rising into the moderate range for tissue texture changes ($\kappa=0.45$) and the substantial range for tenderness assessments ($\kappa=0.68$). Reliability for positional asymmetry in the transverse plane ($\kappa=0.34$) and rotational motion asymmetry ($\kappa=0.20$) improved but remained in the fair range.¹¹ In pelvic anatomical landmark asymmetry, Kmita and Lucas¹² revealed low reliability, with κ ranging from -0.38 to 0.51. Rajendran and Gallagher¹³ studied the interrater reliability on both the pelvis and spine, assessing Mitchell's pelvic diagnostic procedures,¹⁴ and obtained a κ statistic ranging from -0.05 to 0.03. Some of these studies used TART to describe the findings. Lower limb area was also studied by Kmita and Lucas¹² to assess the interrater reliability on the medial malleoli asymmetry test, and they obtained κ levels ranging from -0.05 to 0.49.¹² Overall, the systematic review by Basile et al¹⁵ reported that the levels of diagnostic reliability in osteopathy

were heterogeneous. Bush and Vorro¹⁶ acknowledged that reaching a reliable palpatory diagnosis is a challenging task. Therefore, they suggested that palpatory findings could be correlated with kinematic parameters to objectify them to increase the reliability.

To our knowledge, no studies have addressed interrater reliability of osteopathic sacral palpatory diagnostic tests using all of the TART parameters, despite the frequent occurrence of somatic dysfunctions in this anatomic region (12% of tested patients according to Licciardone et al¹⁷). The aim of this study was to evaluate the interrater reliability among students of an Italian osteopathy school at the end of their curriculum. Our hypothesis was that 3 trained osteopathy students at the end of their curriculum would achieve at least moderate agreement on osteopathic sacral palpatory diagnostic tests.

Methods

This investigation was a prospective interrater reliability study on osteopathic sacral palpatory diagnostic tests performed by 3 raters on 52 subjects. The study was carried out in June 2015 at the Centre pour l'Étude, la Recherche et la Diffusion Ostéopathiques (CERDO) in Rome, Italy.

Raters

Three last-year full-time students with no previous health care degree, all aged 24 years, were trained as raters. Students were selected on a volunteer basis after they requested to engage in the research. No volunteers were excluded.

Consensus Training

Preliminary consensus training followed the training described by Degenhardt et al,¹¹ who reported an increase of interrater reliability after a consensus training period. The training aimed to acquire a uniform method of testing each TART parameter and to reach a common understanding of the data collection form. No pretraining interrater reliability was assessed. The training took place over 3 days, 3 hours per day, for a total

of 9 hours of training. Every palpatory diagnostic test was performed on a prone-positioned peer (ie, one of the raters). An osteopath with 1 year of experience in clinical practice supervised the training.

For the tissue texture abnormality parameter, the raters trained by searching for visually evident signs of cutaneous alterations on the sacroiliac joint and by softly rubbing the second and third finger on the skin over the sacroiliac joint searching for boggiess, thickening, stringiness, ropiness, firmness, increased or decreased temperature, and increased or decreased moisture on a subject in a prone position.

For the asymmetry parameter, the training consisted of the search of 4 landmarks: the 2 extremes of the sacral base and the sacral angles. This maneuver was performed by positioning the second finger of both hands at the extremes of the sacral base and the first fingers on the angles. Holding the 4 fingers on the landmarks, the raters had to define the spatial position of the sacrum as the result of the position of every single landmark. The choice of assessing multiple landmarks was supported by evidence, which showed increased accuracy of the assessment when the test was carried out in conjunction with multiple landmarks rather than a single landmark.¹⁸

For the restriction of motion parameter, the raters trained by executing standard osteopathic mobility tests on the sacrum,¹⁹ testing both unilateral and bilateral extended and flexed sacrum and the 4 types of sacral torsion: forward L (left) on L, R (right) on R, and backward L on R, R on L.

For the tenderness parameter, the raters were trained to hold a pressure of 4 kg/cm² on a scale with their thumb.¹¹ They performed 4 attempts. The first attempt was done while looking at the scale to check when 4 kg was reached, the second one by pressing without looking but listening to another rater saying when the correct pressure was reached. The last 2 attempts were both blinded and without any vocal guidance. Afterwards, the raters were trained to apply pressure along the entire sacroiliac joint, and feedback was asked regarding the tenderness perceived by the peer during the execution.

At the end of practical training, one of the researchers (G.C.) presented and discussed the data collection form.

Subjects

A convenience sample of 52 consecutive osteopathy students at CERDO were recruited on a voluntary basis to participate as tested subjects. Subjects were recruited using a notice posted on the school bulletin board and at the door of classrooms. Inclusion criteria were commitment to be present at the palpatory diagnostic test. Subjects were excluded if they had a history of recent trauma, if they had a ligament or bone lesion, if they had a clinical condition that could have interfered with the execution of the palpatory diagnostic test, if they were unable to lie prone for at least 11 minutes, and if they had a history of lumbar spine surgery. Subjects who were symptomatic in the lumbar and sacroiliac area but did not have severe intensity were included.

Ethical approval was delivered by the didactic committee of CERDO, and informed consent was signed by all the subjects enrolled in this study. Sex, age, height, and weight of the subjects were collected, and body mass index was calculated.

Procedure

The study was carried out in 2 sessions at a 5-day interval. Twenty-six subjects were tested on the first day and the remaining 26 on the second day. A room was set up with 3 identical treatment tables 3 m apart. Three subjects entered the room together, and the 3 raters were randomly assigned to start the examination. Then all subjects undressed, keeping on their undergarments, and laid prone on the treatment table. Their undergarments were lowered to the superior extremities of the intergluteal line. The rotation sequence was also determined randomly, so that each subject was examined by all 3 raters. The rotation was guided by one of the researchers (G.C.). This procedure ensured that the third rater would not have a track of the previous manipulation from the same rater, which could produce a bias influenced by the personal style of touch. All of the raters were blinded to the clinical conditions of the subjects.

The transcription of data on the collection form was executed on separate tables so that each rater was blinded to the findings of the other raters. Blinding was guaranteed by the presence of a curtain between each treatment table, by the instructions given to the subjects not to communicate with raters apart from the “tenderness test” responses, by the presence of curtains between each data collection table that prevented the raters from viewing the sheets, and by the presence of one of the researchers in the assessment room who guaranteed adherence to the blinding protocol. The data collection form consisted of a prefilled form with all of the possible choices among all TART parameters in which the raters marked their findings. All of the raters followed the same protocol of assessment based on the same sequence of tests, as trained in the consensus training.

The raters started by testing the tissue texture abnormality parameter. The test consisted of searching for visually evident signs of cutaneous alterations on the sacroiliac joint and softly rubbing the second and third finger on the skin over the sacroiliac joint searching for boggianness, thickening, or stringiness. The finding was considered positive if at least 1 of the visual or palpatory findings were positive.

The asymmetry parameter consisted of 2 distinct values: the presence or absence of asymmetry and the assessment of all 4 landmark positions (the 2 extremes of the sacral base and the sacral angles), according to the spatial position of the sacrum. The agreement on recognition of all 4 landmark positions was reached if the raters indicated the same position for each landmark. The raters had to find the presence or absence of 1 of 2 possible landmark positional asymmetries, such as “superior” or “inferior” and “anterior” or “posterior.”

For the restriction of motion parameter, the raters moved the sacrum around the middle transverse axis and around both oblique axes. The test was considered positive if at least 1 of the movements around those axes lacked on amplitude.

The test for the tenderness parameter consisted of applying 4 kg/cm² of pressure on the entire length of

the sacroiliac joint with the first finger, asking for feedback on tenderness perceived by the subjects. The test was considered positive if during the pressure the subject replied “yes” to the question “does it hurt?” asked by the rater. The subject was asked to whisper the answer to avoid the possible blinding bias of being heard by the other raters. Raters were also obliged to report on their form the answer received from the subject even in the unlikely circumstance they heard something different being said by the same subject with the previous rater.

According to Ehrenfeuchter and Kappler,²⁰ the diagnosis of somatic dysfunction required the presence of at least 2 of the TART parameters. The somatic dysfunction was defined choosing from 1 of the somatic dysfunctions in which raters trained during the consensus training, respectively: both unilateral and bilateral extended and flexed sacrum and the 4 types of sacral torsion (forward L on L, R on R, and backward L on R, R on L). The choice was achieved by selecting the facilitated direction of motion on the joint mobility test with the condition of a lack of mobility on the opposite direction (ie, a facilitated forward left mobility on an oblique left axis with an associated restricted motion on backward right mobility on a left axis was defined as forward L on L somatic dysfunction). The absence of somatic dysfunction was a possible choice.

Raters were allowed 3 minutes to assess each subject, with one of the researchers notifying raters of the end of the assessment. The raters were also allowed 1 minute to fill in the data collection form after each subject. Subjects were allowed to get up immediately after the third assessment; therefore, every subject stayed in the prone position for a total of 11 minutes. As students completed the final assessment form, 3 new subjects entered the room, and new assessments were executed.

Statistical Analysis

Interrater reliability was evaluated using the Fleiss κ statistic. This method expresses the extent of agreement among 2 or more raters beyond chance, from poor agreement to almost perfect agreement (Table 1).²¹

κ statistics were computed for the assessment of the 5 values derived from the 4 TART parameters and for the detected somatic dysfunctions or the absence of somatic dysfunction.

Results

A total of 52 subjects were enrolled in the study: 20 women (38.5%) and 32 men (61.5%), with a mean (SD) age of 25.9 (7.03) years. No subjects met any of the exclusion criteria. The characteristics of the subjects are summarized in Table 2. Thirty-one subjects (59.6%) had pain in the sacroiliac or lumbar area, but pain intensity varied.

The interrater agreement among the raters was slight or fair, ranging from 0.06 to 0.34. More precisely, the κ statistics were as follows: 0.28 for tissue texture abnormality, 0.29 for asymmetry, 0.06 for the landmark position, 0.32 for the restriction of motion, and 0.34 for tenderness. The overall κ value for the detected somatic dysfunctions was 0.17 (Table 3).

Discussion

This study aimed to assess the interrater reliability of osteopathic sacral palpatory diagnostic tests among students at the end of their curriculum. To our knowledge, no other studies have specifically considered the interrater reliability of osteopathic palpatory diagnostic tests on the sacral region using all 4 TART parameters.

Table 1.
Qualitative Descriptors of Interrater Reliability by Fleiss κ Statistic²¹

κ	Qualitative Descriptors
<0.00	Poor agreement
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

Table 2.
Characteristics of Subjects Evaluated by Raters Using Osteopathic Sacral Palpatory Diagnostic Tests

Characteristic	Value
Sex, No. (%)	
Men	32 (61.5)
Women	20 (28.5)
Age, mean (SD)	25.9 (7.03)
Height, m, mean (SD)	1.7 (0.09)
Weight, kg, mean (SD)	68.7 (14.2)
Body Mass Index, mean (SD)	22.7 (3.6)
Symptomatic in Lumbar Area, No. (%)	31 (59.6)

Previous studies,^{4,11,22} in addition to our results, showed that tenderness had the strongest interrater reliability of the TART parameters. The landmarks position item showed the overall lowest reliability in our results, similar to asymmetry in the study by Degenhardt et al.¹¹ A possible explanation is that the tenderness parameter is the easiest one to achieve by raters from a technical point of view, and it is the only one that evoked an explicit answer from the subjects. The asymmetry parameter, on the other hand, is technically more complex and multifactorial because it requires a judgement on the orientation in the 3 spatial landmark planes. The reliability of the other parameters spans in between.

We conducted our study using students at the end of their curriculum as raters because no consistent improve-

ment on interrater reliability has been found among experienced raters and trained students.²³ However, this finding was not specifically noted in regard to the evaluation of the sacral region. Future studies among experienced practitioners could clarify the possible influence of experience on palpatory diagnostic tests on the sacral region. It is evident that large variability exists in individuals' anatomy and movement of the sacroiliac joint and that clinical manual movement tests (ie, standing flexion test) are unreliable for the sacroiliac joint.²⁴

Compared with the study by Degenhardt et al,¹¹ our raters had a shorter period of training (9 hours vs 24 hours). This fact could explain the lower interrater reliability we observed, even if a different body region was considered. Furthermore, the training was not long enough to justify a recalibration, which did occur in the study by Degenhardt et al.⁴ However, beside the specific consensus training, our 3 students practiced manual tests from the beginning of their education. They received 66 hours of training on the pelvic region during their curriculum, plus the 9 hours of consensus training before the experimental part of the study. The limited experience of the osteopath who supervised the training process could also explain the possible lack of the efficacy of the process. Future comparative studies could give more information on the actual value of training in increasing interrater reliability. To improve training methods, Howell et al²⁵ suggested that a virtual learning environment could be an effective aid to osteopathic medical students in learning palpatory diagnosis. A uniform evaluation of the TART parameters is needed and could influence the educational strategies for osteopathic curricula.

A major limitation of this study is the use of a convenience sample, which is not representative of all patients who receive osteopathic care. We did not plan a subanalysis regarding the differences in interrater reliability among symptomatic and asymptomatic subjects. This choice was supported by Seffinger et al,²³ who found no differences in interrater reliability between the palpatory diagnostic test performed on symptomatic and asymptomatic subjects. Also, in our experience,

Table 3.
Interrater Reliability of Osteopathy Students in Osteopathic Sacral Palpatory Diagnostic Tests

Parameter	K
Tissue texture abnormality	0.28
Asymmetry	0.29
Landmark position	0.06
Restriction of motion	0.32
Tenderness	0.34
Definition of somatic dysfunction	0.17

osteopaths in clinical practice typically assess the sacrum even if it is asymptomatic. To better understand this aspect, additional data might be useful.

A further possible confounding factor could be the amount of time in which subjects laid in the prone position. A prolonged prone position could change the subjects' condition, leading to a different assessment. We kept that amount of time to 11 minutes, but we cannot exclude that this time had an influence on the observed reliability. Another limitation is the lack of representativeness of the raters, who were 3 students from the same school. As shown by Luciani et al,²⁶ significant differences in preparedness are present among students of European osteopathy institutes. Exhaustive results might be achieved by a multicenter study carried out among several osteopathy schools.

The International Federation for Manual/Musculoskeletal Medicine Scientific Committee is highly critical of interrater reliability studies, stating, "The results of these kinds of studies inform us more about the skills and/or the quality of the educational systems of the observers, rather than about the reproducibility of the evaluated tests."^{27(p18)} Hence, they suggest strengthening the training procedures before the observations.²⁷

At present, there is no agreement on how many TART parameters are needed to diagnose somatic dysfunction. The *Glossary of Osteopathic Terminology* states that "any one of [TART] must be present for the diagnosis."¹ We instead adopted the Ehrenfeuchter and Kappler²⁰ model, which states that "The diagnosis of somatic dysfunction requires at least 2 of [TART]." The decision to diagnose the somatic dysfunction by at least 2 TART parameters could have introduced further bias. We did not compare the statistics of the diagnosis based on 1 or 2 TART parameters; repeating our study with a different definition of somatic dysfunction could be of interest. Furthermore, in clinical practice, every diagnosis of somatic dysfunction is achieved by means of triangulation of different information (ie, anamnesis, physical examination, dysfunctions resulting from other body regions, and other kinds of diagnostic tests).

Petersen et al²⁸ suggested that a multiple tests protocol is a "best-evidence diagnostic rule" for the sacroiliac joint. Raters in our study were blinded to this information, and this blinding could have contributed to increased difficulty in obtaining high reliability levels.

There is a growing number of studies, with different findings, that have investigated the reliability of manipulative diagnosis.^{3-8,29-33} Thus, a deeper understanding of the causes and conditions of this variability is needed, as well as the application of multivariate statistical frameworks like the generalizability theory.³⁴ The outcome of such research could be used to design a more focused consensus training, leading to an increase of reliability as suggested by Degenhardt et al.⁴

Conclusion

To our knowledge, this study is the first to examine interrater reliability of osteopathic sacral palpatory diagnostic tests using all of the TART parameters. We found a level of agreement ranging from slight to fair in the assessment of the TART parameters among students at the end of their curriculum. The tenderness parameter was the most reliable one. Our findings are consistent with other interrater reliability studies carried out on different body regions and contribute to an overall heterogeneous level of diagnostic reliability in osteopathy.¹⁵ Studies with a more representative sample of raters and subjects, and focused on causes that could lead to increased variability, are needed to better understand this key element of osteopathic practice.

Acknowledgments

We thank the raters Lola Masi, DO (Italy), and Francesco Lucci, DO (Italy), for their contributions to the study.

Author Contributions

Mr Consorti and Mr Pugliese provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; Mr Consorti, Mr Basile, and Mr Petracca drafted the article or revised it critically for important intellectual content; all authors gave final approval of the version of the article to be published; and all authors agree to be accountable for all aspects of the work in ensuring that questions related to

the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Educational Council on Osteopathic Principles. *Glossary of Osteopathic Terminology*. Chevy Chase, MD: American Association of Colleges of Osteopathic Medicine; 2011.
2. Licciardone JC, Nelson KE, Glonek T, Sleszynski SL, des Anges Crusier. Osteopathic manipulative treatment of somatic dysfunction among patients in the family practice clinic setting: a retrospective analysis. *J Am Osteopath Assoc*. 2005;105(12):537-544.
3. Lucas N, Bogduk N. Diagnostic reliability in osteopathic medicine. *Int J Osteopath Med*. 2011;14(2):43-47.
4. Degenhardt BF, Johnson JC, Snider KT, Snider EJ. Maintenance and improvement of interobserver reliability of osteopathic palpatory tests over a 4-month period. *J Am Osteopath Assoc*. 2010;110(10):579-586.
5. Bengaard K, Bogue RJ, Crow WT. Reliability of diagnosis of somatic dysfunction among osteopathic physicians and medical students. *Osteopath Fam Physician*. 2012;4(1):2-7.
6. Stovall BA, Bae S, Kumar S. Anterior superior iliac spine asymmetry assessment on a novel pelvic model: an investigation of accuracy and reliability. *J Manipulative Physiol Ther*. 2010;33(5):378-385.
7. Stovall BA, Kumar S. Anatomical landmark asymmetry assessment in the lumbar spine and pelvis. *PM R*. 2010;2(1):48-56.
8. Sutton C, Nono L, Johnston RG, Thomson OP. The effects of experience on the inter-reliability of osteopaths to detect changes in posterior superior iliac spine levels using a hidden heel wedge. *J Bodyw Mov Ther*. 2013;17(2):143-150.
9. Upledger JE. The reproducibility of craniosacral examination findings: a statistical analysis. *J Am Osteopath Assoc*. 1977;76(12):890-899.
10. Moran RW, Gibbons P. Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. *J Manipulative Physiol Ther*. 2001;24(3):183-190.
11. 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.
12. 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.
13. Rajendran D, Gallagher D. The assessment of pelvic landmarks using palpation: a reliability study of undergraduate students. *Int J Osteopath Med*. 2010;14(2):57-60. doi:10.1016/j.ijosm.2010.10.005
14. Mitchell JF, Moran P, Pruzzo N. *An Evaluation and Treatment Manual of Osteopathic Muscle Energy Procedures*. Valley Park, MO: Mitchell, Moran and Pruzzo; 1979.
15. Basile F, Scionti R, Petracca M. Diagnostic reliability of osteopathic tests: a systematic review. *Int J Osteopath Med*. 2017;25:21-29.
16. Bush TR, Vorro J. Kinematic measures to objectify head and neck motions in palpatory diagnosis: a pilot study. *J Am Osteopath Assoc*. 2008;108(2):55-62.
17. Licciardone JC, Nelson KE, Glonek T, Sleszynski SL, des Anges Crusier. Osteopathic manipulative treatment of somatic dysfunction among patients in the family practice clinic setting: a retrospective analysis. *J Am Osteopath Assoc*. 2005;105(12):537-544.
18. Snider KT, Snider EJ, Degenhardt BF, Johnson JC, Kribs JW. Palpatory accuracy of lumbar spinous processes using multiple bony landmarks. *J Manipulative Physiol Ther*. 2011;34(5):306-313.
19. Nicholas AS, Nicholas EA. *Atlas of Osteopathic Techniques*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2011.
20. Ehrenfeuchter WC, Kappler RE. Palpatory examination. In: Chila AG, executive ed. *Foundations of Osteopathic Medicine*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2011:403.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
22. Hubka MJ, Phelan SP. Interexaminer reliability of palpation for cervical spine tenderness. *J Manipulative Physiol Ther*. 1994;17(9):591-595.
23. Seffinger MA, Najm WI, Mishra SI, et al. 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.
24. Vleeming A, Schuenke MD, Masi AT, Carreiro JE, Danneels L, Willard FH. The sacroiliac joint: an overview of its anatomy, function and potential clinical implications. *J Anat*. 2012;221(6):537-567.
25. Howell JN, Conatser RR, Williams RL II, Burns JM, Eland DC. Palpatory diagnosis training on the virtual haptic back: performance improvement and user evaluations. *J Am Osteopath Assoc*. 2008;108(1):29-36.
26. Luciani E, van Dun PLS, Esteves JE, et al. Learning Environment, Preparedness and Satisfaction in Osteopathy in Europe: the PreSS study. *PLoS One*. 2015;10(6):e0129904.
27. FIMM Scientific Committee; Patijn J. *Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine: Protocol Formats*. 3rd ed. International Federation for Manual/Musculoskeletal Medicine; 2004.
28. Petersen T, Laslett M, Juhl C. Clinical classification in low back pain: best-evidence diagnostic rules based on systematic reviews. *BMC Musculoskelet Disord*. 2017;18(1):188.
29. 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.
30. Halma KD, Degenhardt BF, Snider KT, Johnson JC, Flaum MS, Bradshaw D. Intraobserver reliability of cranial strain patterns as evaluated by osteopathic physicians: a pilot study. *J Am Osteopath Assoc*. 2008;108(9):493-502.
31. 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.
32. Paulet T, Fryer G. Inter-examiner reliability of palpation for tissue texture abnormality in the thoracic paraspinal region. *Int J Osteopath Med*. 2009;12(3):92-96.
33. Potter L, McCarthy C, Oldham J. Intraexaminer reliability of identifying a dysfunctional segment in the thoracic and lumbar spine. *J Manipulative Physiol Ther*. 2006;29(3):203-207.
34. Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: a liberation of reliability theory. *Br J Stat Psychol*. 1963;16:137-163.

© 2018 American Osteopathic Association