Assessment of Consistency Between the Arm-Fossa Test and Gillet Test: A Pilot Study

Robert Cooperstein MA, DC ^{a,*}, Charles Blum DC ^b, and Elaine C. Cooperstein MS, DC ^c

^a Professor Palmer Center for Chiropractic Research, Palmer Chiropractic College, West, San Jose, CA

^b Research Chair, SOTO – USA, Sparta, NC

^c Private Practice, San Jose, CA

Received 29 June 2014; received in revised form 9 October 2014; accepted 13 October 2014

Key indexing terms:

Sacroiliac joint;
Palpation;
Range of motion,
articular;
Reproducibility of
results;
Validity;
Muscle strength;
Chiropractic

Abstract

Objective: The purpose of this pilot study was to test methods needed to conduct a study with adequate power to investigate consistency between the arm-fossa test (AFT) and the Gillet test.

Methods: A convenience sample of chiropractic college students enrolled in a weekend Sacro-Occipital Technique seminar participated. Each was tested with AFT and sacroiliac orthopedic tests, including the Gillet test. Statistical testing included calculation of κ for consistency of the AFT and Gillet test and their diagnostic efficiency.

Results: This study recruited 14 participants. Important issues arose in gathering and recording data, the standardization of examiner methods, and the flow of participants to examination stations. κ for AFT and Gillet test consistency = 0.55, corresponding to “moderate.”

Conclusion: This pilot suggests that the future study should include a mix of symptomatic and asymptomatic participants; record trichotomous data, where appropriate; use washout periods between diagnostic tests; and refine the selection of orthopedic tests deployed besides the AFT. The preliminary data are consistent with but do not establish due to the very small sample size and experimental design issues, that a positive AFT may be consistent with a negative Gillet test.

© 2015 National University of Health Sciences.

Introduction

The arm-fossa test (AFT), also called the “arm pull-down test,” is a manual muscle test used by practitioners of Sacro-Occipital Technique (SOT), a proprietary chiropractic technique originally developed

by Dr DeJarnette.¹ It tests the ability of a supine patient to maintain the arm in a flexed position, whereas the examiner applies pressure inferiorly and simultaneously makes light contact with a series of contact points on the inguinal ligaments (see Fig 1). The AFT is considered “positive” if the arm “weakens” and “negative” if it does not. Sacro-Occipital Technique practitioners believe that a positive AFT is associated with a complex of findings including sacroiliac hypermobility, whereas a negative AFT is consistent

* Corresponding author. PCCW, 90 East Tasman Drive, San Jose, CA 95134. Tel.: +1 408 944 6009; fax: +1 408 944 6118.
E-mail address: Cooperstein_r@palmer.edu (R. Cooperstein).

Fig 1. The AFT for sacroiliac hypermobility. (Color version of figure appears online.)

with (but does not necessarily identify) sacroiliac hypomobility.² Because a positive AFT is thought to be related to sacroiliac dysfunction, the authors hypothesized that its findings might be consistent with sacroiliac motion tests such as the Gillet test (see Fig 2).³ The Gillet test determines if anatomic landmarks on the sacrum and ilium diverge during flexion of the hip in the standing position. A positive finding is thought to identify sacroiliac fixation: if so, a negative Gillet test would be consistent with (but not necessarily identify) a positive AFT.

The AFT is poorly studied, despite that thousands of SOT practitioners use it every day.⁴ Gatterman et al⁵ recommended that the research community prioritize areas of attention by focusing on procedures that have little supportive evidence but very widespread use. The authors know of only 1 study in the peer-reviewed literature investigating the association of the AFT and Gillet tests: in a 1988 validity study, Leboeuf et al⁶ reported no significant association between a positive AFT and the side of a fixated sacroiliac joint. Because no AFT-negative participants were entered into that study, specificity and sensitivity calculations were precluded. A more thorough investigation of the association of AFT and Gillet test results would include both AFT-positive and AFT-negative participants.

Although the Gillet test has not been demonstrated to be reliable in most studies,⁷ Hungerford et al⁸ showed good reliability for 1 version of the test. Cooperstein et al⁹ suggested that limiting hip flexion in the Gillet test to approximately 30°, rather than the more typical 90°, may enable more accurate detection

Fig 2. The Gillet test for sacroiliac fixation. (Color version of figure appears online.)

of sacroiliac fixation. Besides the validity study of 1988,⁶ Leboeuf^{10,11} and Leboeuf et al^{6,12} reported the following series of results for the AFT: high¹² and low¹⁰ intraexaminer reliability, low interexaminer reliability¹⁰, some validity in correctly distinguishing a correctly treated from an incorrectly treated group of participants⁶, and some validity¹¹ in relationship to lumbopelvic pain (sensitivity, 54%; specificity, 69%).

The primary goal of the present pilot study was to assess the feasibility of conducting a future study with adequate power to assess the consistency of both positive and negative AFT findings with Gillet test findings. Besides these tests, the pilot included a group of other sacroiliac and sacroiliac-related examination procedures that could interact not only with the AFT and Gillet tests but with each other (Table 1). The secondary goal was to determine if there was a trend toward consistency among the included orthopedic tests and agreement among examiners. The qualitative assessments of this study (Table 2) would include each of the following: suitability of a teaching seminar in SOT to serve as a venue for gathering exploratory data, cohesiveness of the team of investigators, degree to which pilot data demonstrated the need to standardize the AFT examiners' protocols, appropriateness of each of the included orthopedic tests, practicality of running this series of orthopedic tests in relatively rapid succession, ease of use of the data entry form, degree

Table 1 Orthopedic Tests Used in This Study

Name of Test	How Performed	Purpose
AFT ¹³	Latissimus dorsi response checked with light contacts on inguinal ligament	Differential diagnosis between an SI joint sprain, hypermobility, or instability and normal or fixated SI joint
Unloaded leg check ¹⁴	Inspection of feet, prone, light pressure	Detect apparent LLI
Compressive leg check ^{15,16}	Inspection of feet, prone, several pounds of pressure	Detect anatomic LLI
Sacral sulcus ¹⁷	Palpation of sacral sulci depth	Sacral rotation and fixation
Sitting PSIS ¹⁸	Palpation of seated PSIS locations	Detect pelvic torsion
Sit/stand PSIS ^{19,20}	Comparison of seated/standing PSIS locations	Indirect test for anatomic LLI
Gillet test ³	Check SI excursion with fingers on PSIS and sacral base while hip flexed	Detect fixation of SI joint

AFT, arm-fossa test; LLI, leg length inequality; PSIS, posterior superior iliac spine; SI, sacroiliac.

to which the goals of the teaching seminar were disrupted for students and instructor, and the willingness of seminar attendees and the instructor to serve as research participants in a future study. The success of the pilot study depended on the degree to which its results established the feasibility of a future follow-up study.

Methods

Subjects

A convenience sample of asymptomatic chiropractic college students enrolled in a weekend SOT seminar at a chiropractic college was recruited. The study was

approved by the Institutional Review Board of the Palmer College of Chiropractic, and each participant provided informed consent. The experimental procedures followed were in accordance with the ethical standards of the Palmer College of Chiropractic Institutional Review Board. The inclusion criteria included being able to lie prone, lie supine, or sit comfortably for up to 3 minutes. Participants were excluded if they declined to be touched by an examiner in the inguinal ligament area. The participants provided demographic data on age, sex, height and weight, and history of past and present sacroiliac and/or low back pain. There were 3 examiners in the study. Examiner 1, the instructor for the seminar, has been teaching SOT for over 32 years. Examiner 2 is an instructor at the college who has practiced and taught SOT procedures for several

Table 2 Summarized Qualitative Outcomes for Pilot Study on Agreement of AFT and Gillet Tests

Item	Outcome
Suitability of a teaching seminar in SOT to serve as a venue for gathering exploratory data	Data acquisition was slower than expected, putting in question the suitability of using the teaching seminar for the future study.
Cohesiveness of the team of investigators	Cohesion was satisfactory.
Practicality of running a series of orthopedic tests in relatively rapid succession	The non-AFT examiner in the study felt rushed keeping up with flow of participants, with several negative consequences: likely errors in data recording, inadequate washout period between individual examination procedures, and probably recall bias in judging results.
Degree to which pilot data demonstrated the AFT examiners would need to better standardize their examination protocols to achieve reliability	The poor interexaminer reliability suggests that standardization of examination methods was inadequate.
Appropriateness of each of the included orthopedic tests	The ad hoc decision to make some of the orthopedic examiner findings trichotomous reduced pilot's utility in weaning inappropriate tests from future study.
The ease of use of the data entry form	Because it was entirely designed for dichotomous examination findings, it was inadequate.
Degree to which the goals of the teaching seminar were disrupted for students and instructor.	The 1.25-h duration of data collection was minimally disruptive; many students found participating in research related to the seminar content added to their understanding as did the instructor.
The willingness of seminar attendees and the instructor to serve as research participants.	All participants and the instructor expressed their willingness to assist in future-related research.

AFT, arm-fossa test; SOT, Sacro-Occipital Technique.

years. Examiner 3 instructs at the college in various sacroiliac orthopedic and pain provocative procedures. Two other individuals served as study coordinators.

Procedures

All of the participants were examined with the AFT by both examiners 1 and 2, half starting with examiner 1 and the other half with examiner 2. Between the conduct of the 2 AFT examinations, each participant underwent a battery of orthopedic and motion tests administered by examiner 3. This battery included unloaded prone leg checking, compressive leg checking, sacral sulcus palpation, sitting posterior superior iliac spine (PSIS) palpation, sit-stand test for anatomic short leg, and the Gillet test. These tests are briefly described and referenced in Table 1. Each AFT test required less than a minute, whereas the battery of tests required approximately 3 minutes. Approximately 3 minutes elapsed between repetitions of the AFT.

Given their central importance to the study, the AFT (Fig 1) and the Gillet test (Fig 2) must be described more fully. The AFT was performed with the participant supine, eyes open. The participant's straight arm was flexed at the shoulder to 90°, with the fist lightly clenched and facing the body. The examiner stood perpendicular to the participant, starting on the participant's right side. The examiner then tested the ability of the participant to respond without delay to the spoken command "hold": the examiner used his left hand to pull the tested right arm caudally against resistance. If the participant could resist the caudal pull on the right arm, it was judged amenable to be used in the AFT. The examiner made light contact with superior-lateral area of the inguinal ligament using the closely approximated tips of digits 2 to 5 of his right hand, whereas the participant was instructed to "hold," as the examiner attempted to pull the right arm inferiorly. The AFT was then repeated, whereas the examiner contacted the inferior-medial area of the right inguinal ligament. The entire examination procedure was then repeated on the left side of the body. The AFT was positive if and when there was a relative delay in arm response upon the examiner's contact with either the superior-lateral or inferior-medial inguinal fossa, on either side of the body. A research assistant using pen and paper recorded whether the particular inguinal fossa contact resulted in a delayed arm response and the side of the body where that occurred.

To perform the Gillet test, the participant stood 12 to 16 inches from the wall without leaning and placed fingertips of both hands on the wall at shoulder level to

maintain balance. A horizontal pencil mark was drawn on the examiner's nail to assist in determining the relative position of the thumbs during performing of the test. The examiner knelt behind the participant, placing 1 thumb on the second sacral tubercle and the other on or near the PSIS, such that the thumbs were horizontally oriented and visualized at the same level. The participant was asked to flex 1 hip and knee, whereas the examiner visualized whether the thumbs diverged vertically on the side of the lifted leg. Hip flexion was limited to approximately 30°. This was repeated for the other side of the body. A positive finding for sacroiliac fixation was recorded if the PSIS on the lifted side failed to drop inferiorly related to the sacral tubercle, as the hip was flexed up to approximately 30°.

Statistics

The chance-corrected level of consistency of the AFT and Gillet tests was analyzed using the κ statistic (Table 3). Diagnostic efficiency was also computed. Statistical calculations were performed using Web-based calculators provided by VassarStats (<http://vassarstats.net/>).

Results

Of 15 participants screened, 14 met the inclusion criteria. The mean age was 27.9 years, and 21.4% were female. All the included participants were asymptomatic, and none reported a relevant prior history of diagnosed sacroiliac injury. The examiner conducting the orthopedic testing other than the AFT found during data acquisition that there ought to have been allowance for trichotomous rather than dichotomous data entry for several of the orthopedic tests (other than the Gillet test, the results of which he continued to feel could be accurately reported dichotomously). For example, the examiner concluded that sacroiliac sulcus depth should be judged to as deep, shallow, or symmetric. Given the small sample size, it became no

Table 3 κ Analysis, AFT, and Gillet Tests

	Positive Gillet	Negative Gillet
Negative AFT (+ for SI fixation)	7 (true-positives)	2 (false-positives)
Positive AFT (– for SI fixation)	1 (false-negatives)	4 (true-negatives)

AFT, arm-fossa test.

longer meaningful to analyze the agreement of test results that had been trichotomously recorded; there were too few data points in the contingency table cells. The qualitative data addressing the feasibility of conducting a future study are described in Table 2.

The 2 AFT examiners did not show interexaminer reliability ($\kappa = -0.13$). Table 3 (based on the findings of examiner 1) summarizes the data used for calculating κ . The diagnostic accuracy, the percentage of examiner ratings that were judged consistent (7 true-positives + 4 true-negatives, divided by the total of 14 participants), was 79%. κ for the agreement of the 2 tests = 0.55 (0.11, 0.99), which was considered “moderate agreement” according to a standard classification scheme.²¹

Discussion

A pilot study is a small version of the main study that is run to test whether the components of the future study are likely to all work together.²² It does not require power calculations to determine an appropriate sample size because one of its aims is to get a sense of the variability of the data to determine sample size in the future study. Although the quantitative data are herein reported, hypothesis testing was secondary to the primary goal of the pilot. Although the study demonstrated “moderate” intertest consistency and high diagnostic accuracy, the results were very “fragile” and must be carefully interpreted.²³ Shifting even 1 data point, given the $n = 14$ sample size, could markedly change the calculated outcomes.

The 2 orthopedic tests emphasized in this study are inversely related: a positive Gillet test is consistent with a negative AFT, and a positive AFT is consistent with a negative Gillet test. Therefore, the study results suggest “consistency rather than agreement.” Although neither of these tests has been irrefutably confirmed by high-quality studies to be reliable and/or valid, it is common practice in the manual therapy research literature to compare orthopedic tests with equivocal evidentiary bases. Triano et al,²⁴ in their authoritative systematic review of the reliability and validity of methods used by chiropractors to determine the site of care, featured a subsection summarizing “Integrated P.A.R.T.S. Studies” that included several examples of cases, where this has been done.²⁵⁻³⁵ These were articles in which the results of a panel of examination techniques, mostly not known to be reliable and/or valid, were combined to guide the decision to provide care. As another example,

there are several articles that, like the present article, supported diagnosing sacroiliac disorders based on the findings of a cluster of tests, although the individual tests were equivocal in their clinical utility.³⁶⁻³⁹ The rationale for this pilot was akin to that of a typical clinical prediction rule study, where a litany of history and physical examination findings—many of which are equivocal or unstudied—is tested for correlation with each other and with a defined clinical outcome. Flynn et al⁴⁰ included 18 sacroiliac/pelvic orthopedic tests in their study identifying patients with low back pain who improve with spinal manipulation; most were not known to be valid or reliable. In their study, the Gillet test was found to be the most reliable of the sacroiliac tests. Unlike Flynn et al,⁴⁰ the present study did not associate test results with a specific clinical outcome.

In comparing the 2 diagnostic tests in this study, it would have been “conceptually” possible to calculate sensitivity and specificity. However, because neither the AFT nor the Gillet test has been found valid, the calculations would have been a function of which test had been arbitrarily defined as the reference standard. For this pilot study, there was no reason to support either test serving as a more credible reference standard than the other. Therefore, there was no rationale for calculating sensitivity and specificity.

According to Evans,⁴¹ the basis of the AFT lies in the fact that the inguinal ligament is the anterior equivalent of the sacroiliac ligament and shares innervation with it. Even light contact with the inguinal ligament, where there is sacroiliac instability, would hypothetically adversely affect the function of the sacroiliac joint, decreasing the patient’s ability to efficiently respond to the instruction to “hold” against inferior ward arm pressure. Such delayed responses have been measured by other authors^{42,43} investigating manual muscle testing as performed by manual therapists.⁴⁴ Some authors have suggested that this delay in response may be related to an inhibited feed forward response.^{45,46} Evans⁴¹ (citing Bathie⁴⁷) suggested that these delayed responses occur, when patients “have lost the control of the pelvis in relationship to the pelvic girdle in relationship to the shoulder girdle. The loss of trunk balance is due to the loss of stability of the latissimus anchor at the pelvis.” According to Rozeboom and Blum,⁴⁸ “The AFT is an evaluation tool developed by DeJarnette to differentiate between SI joint fixation or hypermobility”.

Besides the study of Leboeuf et al⁶ comparing AFT and Gillet test findings, the present authors know of 3 other studies with a similar goal, all of them dissertation projects conducted by chiropractic students.^{41,49,50} The

findings of 2 of these projects^{49,50} were difficult to interpret beyond having been little, if any agreement. The third study, that of Evans⁴¹ and Evans and Yelverton,⁵¹ reported $\kappa = 0.40$ for the AFT and Gillet test. Although their data seem to resemble those of the pilot, the results are not comparable because agreement is defined in different ways. A true-positive in the study of Evan⁴¹ occurred, when a participant was fixed on Gillet testing and positive on the AFT, whereas in the present study, such a participant would have been classified a false-negative (as defined in Table 3).

Each of the 4 cited predecessor studies investigated correlation of a positive AFT and a positive Gillet test. However, a positive AFT has been interpreted by SOT practitioners to include sacroiliac “laxity” and thus predict a “negative” Gillet test for sacroiliac fixation. Leboeuf et al⁶ acknowledged this ambiguity: “A strained sacroiliac joint, as it is devoid of direct muscular support, would perhaps become hypermobile rather hypomobile. The “fixated” side would therefore not necessarily be on the same side as the arm-fossa failure, if the AFT accurately identifies a jarred and/or traumatized sacroiliac joint”. It is not clear why Evan and the other thesis authors expected the Gillet test to be positive in such cases, which he states “involve a chronic hypermobility at the hyaline area of the sacroiliac joint”⁴¹ (p. 43). The 1988 study of Leboeuf et al⁶ reported no significant relationship between the finding of a fixated sacroiliac joint and the side of the “involved fossa.” However, the study of Leboeuf only included participants who had already been found positive for the AFT, precluding comparison with the present study’s findings.

The reason for the “poor” interexaminer reliability for the AFT in this study may lie in the large difference in the examiners’ clinical experience: examiner 1 is a certified SOT practitioner with 3 decades of experience; examiner 2 is less experienced. This might also explain why only examiner 1’s results were concordant with the sacroiliac motion testing. As another possibly contributing factor, the examiners did not rehearse and standardize their AFT test procedures for the limited purposes of this pilot study. The poor interexaminer reliability data suggest that the future study should ensure standardization of examination protocols.

Limitations of the Study

Various logistic issues surfaced during the conduct of this pilot study (Table 2). One of the AFT examiners found the data sheet confusing and feels he may have entered data incorrectly. The data sheets permitted only

dichotomous choices, whereas sometimes, minimal or no left-right differences were detected (in leg checks, sulcus depth palpation, and PSIS levels). Improvising the communication of equivocal results to the data recorder (eg, there is a “slight difference”) resulted in some trichotomous data entries and possible errors in data entry. These ad hoc trichotomous data entries for some of the examination findings precluded calculating the agreement of the multiple orthopedic tests, given the small sample size (although trends could still be seen). The future study, depending on sample size, will exclude certain tests performed in the present study.

In principle, the findings of examiner 3 could have been subject to expectation bias due to having performed 6 orthopedic tests in rapid succession; his examination findings for a test may have been influenced by the results of prior tests. Although he felt that the rapid pace of data collection mitigated potential bias, a future study should be structured to develop a means of eliminating or at least substantially reducing the risk of expectation bias either by using multiple examiners or by allowing for a washout period between the procedures. Likewise, if the future study includes an AFT interexaminer module, each examiner should complete the AFT before participants’ undergoing the sacroiliac battery, the mechanical stresses of which might impact upon the results of a second AFT.

The most obvious limitation of this study is that it is generally inappropriate and possibly unethical²² to extrapolate the results of a small pilot to other clinical settings, not because the sample size is small but because the participants may not be typical of the patient populations to whom its results might be applied. The “moderate” consistency of the AFT with the Gillet test in this pilot study must be interpreted cautiously and should not be presumed to provide evidence of external validity. Under the best of circumstances, a test found reliable or to exhibit findings consistent with other tests may not be considered clinically useful, unless it has also been shown to be valid (ie, accurate in relation to a reference standard).

The participants in this pilot were homogeneous and asymptomatic, not representative of most of the patients who likely receive these tests in clinical circumstances. The rationale for involving a more heterogeneous participant pool in a future study, including symptomatic participants, would be to determine if the orthopedic tests used in this study have greater or lesser consistency among subsets of the participant pool. If symptomatic participants were included in a future study, it would then be possible to compute odds ratios to see for the association of SI dysfunction and pain.

The authors made the methodological decision to conduct AFT before the orthopedic battery because we thought the latter, having so many components, was more likely to influence the AFT than the other way around. Although that decision protected the independence of the AFT from Gillet testing, it came at the expense of introducing an order effect. The costs and benefits of alternative sequences for data acquisition will be carefully considered in the future study.

Future Studies

This pilot study was conducted to inform the design of a future study with a larger sample size, more specific hypotheses, more rigorous masking procedures, more carefully selected sacroiliac orthopedic tests, and more heterogeneous participants. The quantitative data, although of secondary importance, will enable weaning from the design of the future study those orthopedic tests whose results seemed unrelated to the various other tests. The future study may inform whether the AFT merits inclusion in a battery of sacroiliac tests, finds utility as a standalone test with diagnostic value, or neither.

Conclusion

The results of this pilot study identified methodological issues that will need to be addressed in a follow-up study intended to more rigorously study the interaction of the Gillet test and the AFT. The internal consistency seen in the preliminary data suggests but does not confirm the possibility of such interaction. This pilot successfully established the feasibility of conducting a future study, the methods of which will be shaped by the present study.

Funding Sources and Conflicts of Interest

No funding sources were reported for this study. Dr Charles Blum is the Research Chair for Sacro Occipital Technique Organization.

References

- DeJarnette MB. *Sacro Occipital Technic*: 1983. Nebraska City, Nebraska: Self-published; 1983 [287 pp.].
- Cooperstein R. Technique system overview: *Sacro Occipital Technique*. *Chiropr Tech* 1996;8(3):125–31.
- Bergmann T, Peterson DH. *Chiropractic technique*. St. Louis: Elsevier; 2011:35–83.
- Christensen MG. *Job analysis of chiropractic*. A project report, survey analysis and summary of the practice of chiropractic within the United States. Greeley, Colorado: National Board of Chiropractic Examiners; 2005. [208 pp.].
- Gatterman MI, Cooperstein R, Lantz C, Perle SM, Schneider MJ. Rating specific chiropractic technique procedures for common low back conditions. *J Manipulative Physiol Ther* 2001;24(7):449–56.
- Leboeuf C, Jenkins DJ, Smyth RA. Sacro-Occipital Technique: an investigation into the relationship between the arm fossa test and certain examination findings. *J Aust Chiropr Assoc* 1988;18(3):97–9.
- Young M, Cooperstein R. Reliability of the standing hip flexion test: a systematic review. *J Chiropr Educ* 2011;25(1):104.
- Hungerford B, Gilleard W, Lee D. Altered patterns of pelvic bone motion determined in subjects with posterior pelvic pain using skin markers. *Clin Biomech* 2004;19(5):456–64.
- Cooperstein R, Young M, Haneline M. At what angle of hip flexion is the Gillet test the most effective for detecting sacroiliac motion? *J Chiropr Educ* 2011;25(1):76–7.
- Leboeuf C. The reliability of specific Sacro-Occipital Technique diagnostic tests. *J Manipulative Physiol Ther* 1991;14(9):512–7.
- Leboeuf C. The sensitivity and specificity of seven lumbo-pelvic orthopedic tests and the arm fossa test. *J Manipulative Physiol Ther* 1990;13(3):138–43.
- Leboeuf C, Jenkins DJ, Smyth RA. Sacro-Occipital technique: the so-called arm fossa test. Intraexaminer agreement and post-treatment changes. *J Aust Chiropr Assoc* 1988;18(2):67–8.
- DeJarnette B. *Sacro-Occipital Technique*. *Today's Chiropr* 1986:97–8 [115].
- Cooperstein R. Leg checking: why and why not? *DC Tracts* 2012;24(2):4–11.
- Cooperstein R, Morschhauser E, Lisi A, Nick TG. Validity of compressive leg checking in measuring artificial leg-length inequality. *J Manipulative Physiol Ther* 2003;26(9):557–66.
- Cooperstein R, Morschhauser E, Lisi A. Cross-sectional validity of compressive leg checking in measuring artificially created leg length inequality. *J Chiropr Med* 2004;3(3):91–5.
- 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.
- Cooperstein R, Crum E, Morschhauser E, Lisi A. Sitting PSIS positions and prone blocking preferences: a preliminary report. *J Chiropr Educ* 2004;18(1):44–5.
- Cooperstein R. Assessment of sitting-standing pelvic landmarks for anatomical LLI: or Darwin's finches and leg checking. *J Am Chiropr Assoc* 2006;43(2):12–4.
- Bourdillon J, Day E. *Spinal manipulation*. 4th ed. Over Wallop, New Hampshire, England: BAS Printers, Limited; 1987 [250 pp.].
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–74.
- Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol* 2010;10:67.
- Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol* 2014;67(6):622–8.

24. Triano J, Budgell B, Bagnulo A, et al. Review of methods used by chiropractors to determine the site for applying manipulation. *Chiropr Man Ther* 2013;21:36.
25. Leboeuf-Yde C, Kyvik KO. Is it possible to differentiate people with or without low-back pain on the basis of test of lumbopelvic dysfunction? *J Manipulative Physiol Ther* 2000;23(3):160–7.
26. French SD, Green S, Forbes A. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. *J Manipulative Physiol Ther* 2000;23(4):231–8.
27. Qvistgaard E, Rasmussen J, Laetgaard J, Hecksher-Sorensen S, Bliddal H. Intra-observer and inter-observer agreement of the manual examination of the lumbar spine in chronic low-back pain. *Eur Spine J* 2007;16(2):277–82.
28. 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):579–86.
29. Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. *J Manipulative Physiol Ther* 1999;22(6):382–9.
30. 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):203–7.
31. 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):465–73.
32. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. *Phys Ther* 1987;67(10):1526–32.
33. Petersen T, Olsen S, Laslett M, et al. Inter-tester reliability of a new diagnostic classification system for patients with non-specific low back pain. *Aust J Physiother* 2004;50(2):85–94.
34. Rhudy T, Sandefur RM, Burk JM. Interexaminer/intertechnique reliability in spinal subluxation assessment: a multifactorial approach. *Am J Chiropr Med* 1988;1(3):111–4.
35. Keating Jr JC, Bergmann TF, Jacobs GE, Finer BA, Larson K. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. *J Manipulative Physiol Ther* 1990;13(8):463–70.
36. Arab AM, Abdollahi I, Joghataei MT, Golafshani Z, Kazemnejad A. Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. *Man Ther* 2009;14(2):213–21.
37. Cibulka MT, Koldehoff R. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. *J Orthop Sports Phys Ther* 1999;29(2):83–9 [discussion 90–2].
38. Kokmeyer DJ, Van der Wurff P, Aufdemkampe G, Fickenscher TC. The reliability of multitest regimens with sacroiliac pain provocation tests. *J Manipulative Physiol Ther* 2002;25(1):42–8.
39. Robinson HS, Brox JI, Robinson R, Bjelland E, Solem S, Telje T. The reliability of selected motion- and pain provocation tests for the sacroiliac joint. *Man Ther* 2007;12(1):72–9.
40. Flynn T, Fritz J, Whitman J, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine* 2002;27(24):2835–43.
41. Evans AC. The validity and specificity of the arm fossa test. [dissertation]Johannesburg: Doornfontein Campus; 2008.
42. Caruso W, Leisman G. A force/displacement analysis of muscle testing. *Percept Mot Skills* 2000;91(2):683–92.
43. Caruso W, Leisman G. The clinical utility of force/displacement analysis of muscle testing in applied kinesiology. *Int J Neurosci* 2001;106(3–4):147–57.
44. Cuthbert SC, Goodheart Jr GJ. On the reliability and validity of manual muscle testing: a literature review. *Chiropr Osteopath* 2007;15:4.
45. Silfies SP, Mehta R, Smith SS, Karduna AR. Differences in feedforward trunk muscle activity in subgroups of patients with mechanical low back pain. *Arch Phys Med Rehabil* 2009;90(7):1159–69.
46. Marshall P, Murphy B. The effect of sacroiliac joint manipulation on feed-forward activation times of the deep abdominal musculature. *J Manipulative Physiol Ther* 2006;29(3):196–202.
47. Bathie R. SORSI SOT, manual. 1st ed. New York: Sacro Occipital Research Society International; 1996.
48. Rozeboom D, Blum CL. Sacral block technic: balancing sacrospinal function: a case report. *World Federation of Chiropractic Proceedings of 10th Biennial Congress, Montreal Canada, Apr 30-May2, 2009. Toronto: The Federation; 2009. p. 222–3.*
49. Holliman M, Holliman M, Patterson L, Norris D, Doane-Reynolds D. Identifying the site of sacroiliac dysfunction: an exploration of conflicting views. [dissertation]Chesterfield, MO: Logan College of Chiropractic; 1995.
50. Ganz N, Jarrett TR, Oblack D, Taylor N. The sacro-iliac joint: an investigation into the relationship between Gillet's test and Sacro-Occipital Technique category II evaluation. [dissertation]Chesterfield, MO: Logan College of Chiropractic; 1995.
51. Evans AC, Yelverton C. The validity and specificity of the arm fossa test. *World Federation of Chiropractic's 12th Biennial Congress. Durban South Africa: Doornfontein Campus; 2013. p. 83.*