The effect of training on the inter-examiner and intra-examiner reliability of static palpation for assessing pelvic landmarks

Hayley Claire McPherson BSc, MHSc

School of Health Sciences

Faculty of Human Development

Victoria University

Principal Supervisor : Gary Fryer BAppSc, ND

Lecturer, School of Health Sciences

Faculty of Human Development

Victoria University

The effect of training on the inter-examiner and intra-examiner reliability of static palpation for assessing pelvic landmarks

General Information

The Journal of Osteopathic Medicine (J Ost Med) has a continuing call for written submissions in the following categories:

Letters to the Editor. As is common in biomedical journals the editorial board welcomes critical response to any aspect of the J Ost Med. In particular, letters that point out deficiencies and that add to, or further clarify points made in a recently published work, are welcomed. The Editorial Board reserves the right to offer authors of papers the right of rebuttal, which may be published alongside the letter.

Authors should note that the following categories are subject to peer review. Author names and institutions will be blinded from the referees. The referees will be selected from a panel according to their area of expertise. Referees comments will be returned to the authors and will include constructive recommendations for changes as deemed necessary.

Research Reports and Reviews. These should be either i) reports of new findings related to osteopathic medicine that are supported by research evidence. These should be original, previously unpublished works. The report will normally be divided into the following sections: abstract, introduction, materials and methods, results, discussion, conclusion, references. or ii) critical or systematic review that seeks to summarise or draw conclusions from the established literature on a topic relevant to osteopathic medicine.

Short review. The drawing together of present knowledge in a subject area, in order to provide a background for the reader not currently versed in the literature of a particular topic. Shorter in length than, and not intended to be as comprehensive as that of the literature review paper. With more emphasis on outlining areas of deficit in the current literature that warrant further investigation.

Research note. Findings of interest arising from a larger study but not the primary aim of the research endeavor, for example short experiments aimed at establishing the reliability of new equipment used in the primary experiment or other incidental findings of interest, arising from, but not the topic of the primary research. Including further clarification of an experimental protocol after addition of further controls, or statistical reassessment of raw data.

Preliminary findings. Presentation of results from pilot studies which may establish a solid basis for further investigations. Format similar to original research report but with more emphasis in discussion of future studies and hypotheses arising from pilot study.

Contributions. Include articles that do not fit into the above criteria as original research. Includes commentary and essays especially in regards to history, philosophy, professional, educational, clinical, ethical, political and legal aspects of osteopathic medicine.

Clinical Practice. Authors are encouraged to submit papers in one of the following formats: Case Report, Case Problem, and Evidence in Practice.

Case Reports usually document the management of one patient, with an emphasis on presentations that are unusual, rare or where there was an unexpected response to treatment eg.an unexpected side effect or adverse reaction. Authors may also wish to present a case series where multiple occurrences of a similar phenomenon are documented. Preference will be given to reports that are prospective in their planning and utilise Single System Designs, including objective measures.

The aim of the **Case Problem** is to provide a more thorough discussion of the differential diagnosis of a clinical problem. The emphasis is on the clinical reasoning and logic employed in the diagnostic process.

The purpose of the **Evidence in Practice** report is to provide an account of the application of the recognised Evidence Based Medicine process to a real clinical problem. The paper should be written with reference to each of the following five steps: 1. Developing an answerable clinical question. 2. The processes employed in searching the literature for evidence. 3. The appraisal of evidence for usefulness and applicability. 4. Integrating the critical appraisal with existing clinical expertise and with the patient's unique biology, values, and circumstances. 5. Reflect on the process (steps 1 – 4), evaluating effectiveness, and identifying deficiencies. Written guidelines for the Clinical Practice section are available by email (journal@paradise.net.nz).

MANUSCRIPT PREPARATION

Submission. Wherever possible manuscripts should be submitted via electronic mail to journal@paradise.net.nz. Manuscripts should be double spaced, and prepared with a margin of at least 25mm. Include a title page noting all author names, qualifications, and author contact address and institutional affiliation. Start each section on a separate page. All pages should be numbered including the title page. Preference is Microsoft Word for Windows or Macintosh formats. Use a sans serif font (Arial, Times New Roman or similar of size 10 or 12point). Include a brief cover letter indicating the section of the journal the manuscript is intended for. Any hardcopy submission must be accompanied by an electronic file saved on a clearly labeled 3.5inch diskette.

The Editors
Journal of Osteopathic Medicine
808/109 Pitt Street, Sydney
NSW 2000
Australia

Copyright. In order to facilitate dissemination of material, copyright must be transferred to the publisher. Please ensure you include a photocopy of the Transfer of Copyright form (provided) with all submissions of written material.

Letters to the Editor. All letters to the editor should be submitted via electronic mail to journal@paradise.net.nz

Units of Measurement. Reporting of data should be in the International System of Units (SI). Most biomedical journals have adopted the metric system as standard.

TEXT PAGES

Title Page. For papers to be submitted for peer review two title pages should be provided. The first should carry just the title of the paper and no information that might identify the author or institution. The second title page should carry (a) a concise informative title; (b) first, middle and last name of each author, with academic qualifications (if applicable); (c) institutional affiliation; (d) disclaimers, if any; (e) name and mailing address, e-mail address, phone and fax numbers; (f) source(s) of support in the form of funding and/or equipment.

Abstract and Key Words. From 2004, volume 7 number 1 onwards manuscripts for both qualitative and quantitative research approaches should be accompanied by a structured abstract. Commentaries and Essays may continue to use text based abstracts of no more than 150 words. All original articles should include the following headings in the abstract as appropriate: Background, Objective, Design, Setting, Methods, Subjects, Results, and Conclusions. As an absolute minimum: Objectives, Methods, Results, and Conclusions must be provided for all original articles. Abstracts for reviews of the literature (in particular systematic reviews and meta-analysis) should include the following headings as appropriate: Objectives, Data Sources, Study Selection, Data Extraction, Data Synthesis, Conclusions. Abstracts for Case Studies should include the following headings as appropriate: Background, Objectives, Clinical Features, Intervention and Outcomes, Conclusions.

Below the abstract please provide a list of 3 to 10 key indexing terms to be published with the abstract. Wherever possible authors should use National Library of Medicine's Medical Subject Headings (MeSH) as indexing terms. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts. A database of MeSH terms and further information is available online at http://www.nlm.nih.gov/mesh/meshhome.html

Text. The text of observational and experimental articles is usually, but not necessarily, divided into sections with the headings: introduction, methods, results and discussion. Long articles may need sub-headings with some sections to clarify their content, especially the results and discussion sections. Other types of articles such as essays should be subdivided as appropriate to enhance clarity and ease of reading.

Table Pages. Type each table on a separate sheet; remember to double-space all data. If applicable, identify statistical measures of variation, such as standard deviation and standard error of mean. If data is used from another published or unpublished source, obtain permission and acknowledge in the legend. Using arabic numerals, number each table consecutively (in the order in which they were listed in the text in parentheses) and supply a brief title to appear at the top of the table above a horizontal line; place any necessary explanatory matter in footnotes at the bottom of the table. Do not submit tables as photographs; do not use any vertical lines within the table.

Diagrams and figures. Legends for figures should be typed double spaced starting on a separate page after any Tables. All figures must be referred to in the text and identified with Arabic numerals in parentheses (eg: Figure 1) Any symbols used in the figure must be identified and explained in the legend. Figures and illustrations must be of professional quality. All photographic images should ideally be submitted as separate digital files (.tiff, or .jpg), however clean camera ready artwork will also be accepted. Any hand drawn illustrations should be of professional quality. If a figure has been previously published elsewhere, the original source must be acknowledged in the legend and written permission from the copyright holder must accompany submission to the J Ost Med.

The text of original research for a quantitative or qualitative study is typically subdivided into the following sections:

Introduction. State the purpose of the article. Summarise the rationale for the study or observation. Give only strictly pertinent references and do not review the subject extensively. Do not include data or conclusions from the work being reported.

Materials and Methods. Describe your selection of observational or experimental subjects (including controls). Identify the methods, apparatus (manufacturer's name and address in parenthesis) and procedures in sufficient detail to allow workers to reproduce the results. Give references and brief descriptions for methods that have been published but are not well known; describe new methods and evaluate limitations.

Indicate whether procedures followed were in accordance with the ethical standards of the institution or regional committee responsible for ethical standards. Do not use patient names or initials. Take care to mask the identity of any subjects in illustrative material.

Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Discuss eligibility of experimental subjects. Give details about randomisation. Describe the methods for, and success of, any blinding of observations. Report losses to observation (such as dropouts from the study). References to statistical methods should be to standard works (with pages stated). Specify any computer software used in analysis of data. Restrict tables and figures to those needed to explain the argument of the paper and to assess its support. Use graphs as an alternative to tables with any entries; do not duplicate data in graphs and tables. Define statistical terms, abbreviations and most symbols.

Results. Present results in logical sequence in the text, tables and illustrations. Do not repeat in the text all the data in the tables or illustrations. Emphasise or summarise only important observations.

Discussion. Emphasise the new and important aspects of the study and the conclusions that follow from them. Do not repeat in detail data or other material given in the introduction or the results section. Include implications of the findings and their limitations, include implications and directions for future research. Relate the observations to other relevant studies. Link the conclusion with the goals of the study, but avoid unqualified statements and conclusions not completely supported by your data. State new hypothesis when warranted, but clearly label them as such. Recommendations, when appropriate, may be included.

Acknowledgments. In the appendix one or more statements should specify (a) contributions that need acknowledging, but do not justify authorship (b) acknowledgments of technical support (c) acknowledgments of financial and material support, specifying the nature of the support. Persons named in this section must have given their permission to be named. Authors are responsible for obtaining written permission from those acknowledged by name since readers may infer their endorsement of the data and conclusions.

References. Number references consecutively in the order in which they are first mentioned in the text. Identify references in text, tables and legends by numerals as superscripts. References cited only in tables or in legends to figures should be numbered in accordance with a sequence established by the first identification in the text of the particular table or illustration.

Try to avoid using abstracts as references; 'unpublished observations' and 'personal communications' may not be used as references, although references to written, not oral, communications may be inserted. Include among the references papers accepted but not yet published; designate the journal; and add 'in press' (in parentheses). Information from manuscripts submitted but not yet accepted should be cited in the text as 'unpublished observations' (in parentheses).

When a previously cited reference is used again, it is designated in the text (as a superscript) by the number originally assigned to it by it's first use. Do not assign it another number or list it again in the reference list as 'op. cit.'.

When using an author's name in the text the superscript reference must follow immediately after the name eg.:Smith[7] states that...

When adjacent to punctuation, superscripts are always placed after the punctuation eg.: ...Brown and co-workers,[8] using a new method... or at end of sentence: ...total health of the patient.[9]

References must be verified by the author(s) against the original document. Examples of correct forms of referencing are given below. It is the authors' responsibility to ensure that reference titles are accurate and provided in full.

Journals

Standard journal article. List all authors when six or less; when seven or more, list only first three and add 'and co-workers'.

1. Russel R, Groves P, Taub N, O'Dowd J, Reynolds F. Assessing long term backache after childbirth. British Medical Journal. 1993;306:1299-1303.

Books

Personal author(s)

1. Mitchell FL, Jr. The Muscle Energy Manual. Volume 1. East Lansing, Mich: MET Press;1995:166-167.

2. Travell JG, Simons DG. Myofascial Pain and Dysfunction: The Trigger Point Manual. Baltimore: Williams & Wilkins Co.;1983.

Editor, compiler, chairman as author

3. Ward RC. Ed. Foundations for Osteopathic Medicine. Baltimore: Williams & Wilkins; 1997:1127.

Chapter in a book

Knapp ME. Massage. In: Kottke FJ, Lehmann JF, eds. Krusen's Handbook of Physical Medicine and Rehabilitation. 4th ed. Philadelphia, Penn: WB Saunders; 1990:433-435.

Symposium

4. Aston-Jones G, Valentino R. Brain noradrenergic neurons, nociception and stress: basic mechanisms and clinical implications. In: Willard FH, Patterson MM, eds. Nociception and the Neuroendocrine-Immune Connection, 1992 International Symposium. Indianapolis,Ind: American Academy of Osteopathy; 1992:107-132.

ABSTRACT

Objective: To investigate the effect of training on the inter-examiner and intra-examiner reliability of static palpation of pelvic landmarks.

Methods: Two groups of final year osteopathic students (N = 10) examined ten asymptomatic female participants for symmetry of pelvic landmarks. One group of examiners (n = 5) attended two training sessions to standardise examination findings whereas those in the untrained group (n = 5) did not. Three separate examinations of the anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), medial malleoli and sacral inferior lateral angle (SILA) were performed on every landmark on every participant by all examiners.

Results: The trained group produced slightly higher intra-examiner reliability for the ASIS, PSIS and medial malleoli, and slightly higher inter-examiner reliability for the ASIS and SILA. Generalised Kappa (Kg) scores indicated fair inter-examiner agreement for palpation of the ASIS (Kg = 0.24) and medial malleoli (Kg = 0.31) and slight agreement for the PSIS (Kg = 0.08) and SILA (0.04) in the trained group. The mean level of intra-examiner agreement was substantial for the ASIS (Kg = 0.65) and medial malleoli (Kg = 0.68), moderate for the PSIS (Kg = 0.54) and slight for the SILA (anterior posterior reading, Kg = 0.076) in the trained group.

Conclusions: Intra-examiner reliability was higher than inter-examiner reliability. The training sessions produced a marginal increase in both inter-examiner and intra-examiner reliability, but agreement was still less than acceptable for a clinical test.

Keywords: reliability, palpation, pelvic, osteopathy

INTRODUCTION

Palpatory diagnosis is claimed to play a vital role in the process of clinical problem solving for osteopaths. This clinical art form is used extensively within the manual medicine professions to assess the neuromusculoskeletal system, via the use of touch related neurosensory feedback.[1] Manual medicine practitioners place particular emphasis on palpatory analysis to both examine and treat somatic dysfunctions, as well as to evaluate the effects of manual treatment on the body.

Many authors of osteopathic texts advocate the use of palpation to assess for changes in tissue texture and bony position as well as to detect changes such as pulsations and masses.[2] The use of palpation for osteopathic structural diagnosis often involves the detection of more subtle changes in the tissues and requires considerable skill concentration and practice. Osteopathic authors propose that skilled practitioners should use palpation to detect structural asymmetry, as well as differences in both the quality and total range of movement at a joint.[3,4,5] Other manual therapy disciplines also use palpatory diagnosis as a fundamental part of the clinical problem solving process. Despite the proposed importance of static palpation in the detection of structural asymmetry, very few studies have been conducted in this area and the reliability of palpatory diagnosis is yet to be verified.

Although many osteopathic authors have advocated the use of palpation of pelvic landmarks as a key diagnostic test for determining sacroiliac joint dysfunction, there has been little research to determine the frequency with which osteopaths use these methods in practice. Recently Peace and Fryer[2] surveyed the entire Australian profession for tests used to determine sacroiliac joint dysfunction and although the response rate was relatively low (30% or 168 respondents) it appeared that Australian osteopaths heavily relied on static palpation. The most frequently examined landmarks were the PSIS (94%) and ASIS (89%) while others included the SILA (69%) and medial malleoli (65%).

Haas[6] conducted a review of forty-five original articles in the peer-reviewed chiropractic literature addressing examiner reliability. Of these, only five studies examined the reliability of static palpation for pain, joint position and muscle tonicity, and only two of these had findings that were substantiated by the statistical analyses. This author concluded there were no studies to adequately confirm or refute the reliability of static palpation and recommended further research in this area. Haas [6] also noted a high incidence of flawed methodological and statistical analytic procedures among the reliability studies, because only ten of the 45 articles reviewed were claimed to have adequately supported conclusions.

A multitude of factors may lead to inter-examiner inconsistencies, such as patient or examiner expectations and clinical diagnostic skills, regardless of the discipline involved.[1] It is possible that clinical experience plays a role in improving palpatory diagnostic skills, with the assumption that more experienced practitioners will produce greater inter-examiner reliability. This assumption has not been supported by several studies. Mior *et al.*[7] studied the role of experience in clinical accuracy on sacroiliac motion palpation tests. The researchers examined the inter-examiner and intra-examiner reliability of motion palpation to test the mobility of the sacroiliac joints, comparing a group of chiropractic students with that of a group of chiropractors with over five years experience. No significant differences were found between the two groups and inter-examiner reliability was poor for both the students (K= 0.00 to 0.30) and the experienced clinicians (K= 0.00 to 0.16).[7]

Studies investigating the reliability of static palpation often include a variable level of training for the examiners in order to maximise the level of both inter and intra-examiner reliability. Gerwin *et al.*[8] investigated the inter-examiner reliability of myofascial trigger point examination and initially found poor inter-examiner reliability. In a second attempt using the same examiners, these authors included a period of training to standardise examination findings, which established acceptable inter-examiner reliability for the diagnosis of myofascial trigger points.[8]

The inter-examiner and intra-examiner reliability for assessing sacroiliac landmarks using palpation and observation was investigated by O'Haire *et al.*[9] They found intra-examiner reliability was generally greater than inter-examiner reliability, which concurs with previous research in this field.[10] Palpation of the posterior superior iliac spine (PSIS), sacral inferior lateral angle (SILA) and sacral sulcus (SS) produced intra-examiner reliability ranging from substantial to less than chance (Kg = -0.05 to 0.69) with only slight inter-examiner reliability (Kg= 0.04 to 0.08).[9] Inter-examiner agreement may have increased by standardising exactly what constituted a difference in the symmetry of anatomical landmarks, rather than leaving it to the examiners discretion. Although these researchers conducted a pre-study training session, it is unclear as to whether standardisation of findings occurred in the training session and the degree to which examiners constituted an asymmetrical finding may have been inconsistent.

McConnell *et al.*[11] investigated the agreement of neuro-musculoskeletal examination findings by a group of osteopathic physicians using their own procedures. Six examiners were required to assess patients with acute spinal pain, however only the primary examiner knew the patients' history. The study design was poorly controlled throughout with no limitations being placed on the spinal areas to be assessed or the examination procedures to be used. Under these conditions it would have been extremely difficult to achieve any level of agreement. McConnell *et al.*[11] reported low inter-examiner agreement on the intensity of the findings and segmental location. They concluded that if high levels of inter-examiner agreement are to be achieved, the examiners must first agree upon the areas to be examined, the test procedure to be used, the method of quantifying the intensity of the findings and the method of recording.

Keating *et al.*[10] investigated the inter-examiner reliability of noninvasive methods of assessing lumbar spinal segments. They reported fair to moderate agreement beyond chance for palpation of tenderness over paraspinal soft tissues and osseous structures (K = 0.30 to 0.48). Active and passive spinal motion palpation and static palpation for misalignment produced little agreement beyond chance (K = 0.0 to 0.1).[10] However, of these six procedures, two were performed in the seated position while four were performed prone and examiners were free to repeat examinations. These changes in position may have lead to examiner inconsistencies, thereby decreasing the chance of establishing reliability. Although it does not reflect clinical practice, it is likely that reliability will improve if the subject's are instructed to remain as still as possible.

Haas[12] produced guidelines for authors conducting reliability research, in an attempt to review and promote the appropriate use of statistics for reliability and interpretation of results. He argued that Kappa was the statistic of choice for nominal data, overall Kappa should be used to average Kappa values, reliability should be reported segment by segment (not collapsed to give regional reliability), a representative sample of subjects should be used and both examiners and subjects should be adequately blinded.[12]

Research supporting the reliability of static palpation is scarce, while determining methods used to improve agreement are even less common. In addition to the uncertainty of the reliability of palpation for symmetrical landmarks, the validity and usefulness of such findings is questionable. However, if reasonable reliability for this measurement cannot be demonstrated, the question of validity is no longer necessary. Given the apparently widespread use of static palpation in the osteopathic profession, research of these issues is urgently required. The aim of this study was to investigate the effect of training on the inter-examiner and intra-examiner reliability of static palpation of pelvic landmarks.

METHODS

Participants

All participants were recruited from Victoria University and consisted of a group of ten final year osteopathic students as examiners and ten female volunteers as participants. All the participants were female in order to reduce the possibility of the examiners remembering the results. All participants provided written informed consent and the study was approved by the Victoria University Human Research Ethics Committee. The participants were provided with identical black Lycra shorts in order to reduce any identifiable characteristics. Participants were excluded from the study if they had any visibly identifiable characteristics (e.g. tattoos or skin lesions), could not lie still for the duration of the testing procedure (e.g. itching, pain) or if they experienced any difficulties lying either supine (e.g. pregnancy) or prone (e.g. pain). A total of eleven participants volunteered for the study and one was excluded due to the presence of a tattoo on their lower back.

Training for Examiners

The examiners were divided into two groups of five. Group 1 attended two one-hour training sessions in which examiners practiced palpating the pelvic landmarks with an experienced osteopath (GF) and discussed their consistency in analysing the landmarks. The examiners agreed on exactly how much of a difference between the landmarks constituted an asymmetrical finding, in order to maximise agreement between examiners. Group 2 did not attend a training session in order to observe the effect of training on the levels of agreement.

Experimental procedure:

The examination room consisted of ten plinths placed in a circle. A table was placed at the head of each plinth which held the results cards and a box for their collection. Participants initially lay prone on the plinths for examination of the posterior pelvic landmarks (PSIS, SILA), with identical grey sheets placed over their upper bodies. Examiners had up to one minute to palpate each landmark and record their finding on the results card and post it in the participant's collection box. Examiners left the room after all ten examiners completed palpating all ten participants. They then participated in distracting tasks (e.g. visual memory games) in order to reduce the chance of memorising previous findings. Examiners re-entered the room and were allocated to a new starting position. This process continued until each examiner had palpated each landmark on every participant for a total of three times. The process was identical for palpation of the anterior landmarks. The examiners left the room and all participants moved into the supine position with the sheets placed comfortably over their heads so as to allow adequate breathing space. Participants were encouraged to lie as still as possible throughout the duration of the testing procedure.

Procedure for assessment of anatomical landmarks

The examination procedure carried out by the trained examiners began by standing on the side of the plinth corresponding with their dominant eye (e.g. left-eye dominant examiners always stand on the left hand side of each participant) as recommended by osteopathic authors.[3,4] The following landmarks were examined:

1. PSIS
2. SILA
3. ASIS
4. Medial Malleoli

Examination of the PSIS and SILA was performed with the participants lying prone and examination of the ASIS and medial malleoli was performed with the participants lying supine. When the anatomical landmark had been identified via both observation and palpation, the examiners aligned their dominant eye central to the appropriate anatomical landmark. The examiners assessed and recorded the existence of any apparent asymmetry between the right and left anatomical landmarks as R=L, R>L or L>R.

Examination of the PSIS first involved visual identification of the skin dimple, indicating the location of the attachment of the deep fascia.[4] Examiners then positioned their hands over the participant's iliac crests and placed their thumbs on the inferior slope of the PSIS to assess the level of their thumbs in the horizontal plane.[3]

Examiners identified the SILA by palpating down the sacral crests to the sacral hiatus and then moving their thumbs laterally approximately 1.5-2cm.[3] The position of the SILA was assessed firstly in the posterior/anterior orientation by placing the thumb pads on the posterior surface of the SILA for evaluation in the coronal plane.[3] The SILA was then assessed in the superior/inferior orientation by sliding the thumbs under the inferior aspect of the SILA for evaluation in the horizontal plane.[3]

In order to examine the ASIS, examiners placed the palms of their hands over each ASIS to identify the landmark. The examiners then placed their thumb pads on the inferior aspect of the ASIS for assessment against the horizontal plane.[3,4]

To inspect the medial malleoli the examiners stood at the foot of the table and placed their thumbs under the distal ledge of the landmark to evaluate the relative positions of the malleoli against the horizontal plane.[4]

Recording of results

The table beside each plinth contained a set of results cards for recording each participant's findings. To record inter-examiner reliability findings a single card was provided to record the results of one participant by one examiner, hence a set of ten result cards were placed on each plinth to be completed by all ten examiners. The information on the result cards consisted of the name of the anatomical landmark being palpated, the examiners initials and the result. The examiners then circled one of the following possible results corresponding to the position of the landmark: right equals left ($R = L$), right higher than left ($R > L$) or left higher than right ($L > R$). Examiners palpated each participant three times, hence each examiner recorded their results on three separate cards for each landmark.

Statistical Analysis

The data was analysed using the Generalised Kappa (Kg) statistic, which evaluated concordance between examiners (testing inter-examiner reliability) and within examiners (testing intra-examiner reliability). $Kg = (Po - Pe)/(1 - Pe)$ where Po is the proportion of observed agreement between examiners, and Pe is the proportion of agreement achieved by chance alone. Kg is therefore the proportion of observed agreement above chance divided by the maximum possible agreement above chance for perfect agreement between examiners.[12] For perfect agreement Kg = 1, for chance agreement Kg = 0 and when chance agreement is greater than observed agreement Kg is negative. The guidelines proposed by Landis and Koch[13] were used to interpret the kappa values. These guidelines advise the use of values for agreement between 0 and 1 (Table 1). The null standard error, SEo was used to test for a significant difference from chance agreement. In a study of this type, the Kappa values are of greater importance when analysing the results than the level of significance; hence a sample size of ten was sufficient.

INSERT TABLE 1

RESULTS

Intra-examiner Reliability:

The intra-examiner agreement for static palpation of the ASIS ranged from moderate to substantial agreement in the trained group (Kg = 0.54 to 0.80; mean Kg = 0.65) and from less than chance to fair in the untrained group (Kg = -0.01 to 0.40; mean Kg = 0.19), while for the PSIS agreement ranged from slight to almost perfect in the trained group (Kg = 0.12 to 0.83; mean Kg = 0.54) and fair to substantial in the untrained group (Kg = 0.22 to 0.76; mean Kg = 0.49). Assessment of the medial malleoli revealed fair to substantial agreement in the trained group (kg = 0.26 to 0.75; mean Kg = 0.68) and fair to almost perfect agreement in the untrained group (Kg = 0.37 to 0.86; mean Kg = 0.59). Agreement for the SILA was less than chance to fair agreement in the trained group for the anterior-posterior readings (Kg = -0.40 to 0.39; mean Kg = 0.076) and less than chance to substantial in the untrained group (Kg = -0.08 to 0.64; mean Kg = 0.20). For the superior-inferior readings, agreement for the SILA was less than chance to substantial in the trained group (kg = -0.10 to 0.64; mean Kg = 0.20) and less than chance to fair in the untrained group (Kg = -0.17 to 0.28; mean 0.38). Table 2 provided an overview of these results.

INSERT TABLE 2

Inter-examiner Reliability:

The inter-examiner agreement for the ASIS was fair in the trained group (Kg = 0.24) and less than chance in the untrained group (Kg = -0.01), while the agreement for the PSIS was slight in both the trained (Kg = 0.08) and untrained (Kg = 0.15) groups. The agreement for the medial malleoli was fair in both the trained (Kg = 0.31) and untrained (Kg = 0.28) groups. The agreement for the SILA was slight in the trained group for both anterior-posterior (Kg = 0.04) and superior-inferior readings (Kg = 0.04) and less than chance for both anterior-posterior (Kg = -0.01) and superior-inferior (Kg = -0.01) readings in the untrained group. Table 3 provided an overview of these results

INSERT TABLE 3

DISCUSSION

Analysis of the intra-examiner reliability for the palpation of static pelvic asymmetry revealed a variety of levels of agreement and it was evident that some landmarks were more reliably palpated than others. This large range of agreement did not extend to the inter-examiner reliability to the same degree, which ranged from fair to less than chance for both groups of examiners, however the trend towards greater agreement for particular landmarks was evident.

It is important to note that the inter-examiner agreement is of greater importance than intra-examiner agreement when interpreting reliability. Similarly the Kappa value is of greater importance than the significance level when interpreting the results. This study demonstrated a fair level of inter-examiner reliability of palpation of the ASIS (Kg = 0.24) and medial malleoli (Kg = 0.31), but found only slight inter-examiner reliability for palpation of the PSIS (Kg = 0.08) and SILA (Kg = 0.04) in the trained group. Intra-examiner reliability was generally higher than inter-examiner reliability, ranging from slight to almost perfect agreement. In addition, the results revealed that those in the trained group demonstrated a marginal improvement in the intra-examiner reliability for three of the four anatomical landmarks palpated, while for only two of the four landmarks an improvement in inter-examiner reliability was observed following training.

The effect of training on the levels of agreement was most obvious for the inter-examiner reliability of the ASIS, where the agreement was fair (Kg = 0.24) in the trained group and less than chance (Kg = -0.01) in the untrained groups. Similarly the intra-examiner agreement for the ASIS was moderate to substantial (Kg = 0.54 to 0.80) in the trained group and less than chance to fair (Kg = -0.01 to 0.40) in the untrained group.

The training session had a marginal effect on the intra-examiner reliability for the PSIS, with agreement ranging from slight to almost perfect (Kg = 0.12 to 0.83) in the trained group and fair to substantial (Kg = 0.22 to 0.76) in the untrained group. Training also appeared to have little effect on the inter-examiner reliability of both the PSIS and medial malleoli. The inter-examiner reliability of the malleoli was fair in both the trained (Kg = 0.31) and untrained group (Kg = 0.28), while the intra-examiner reliability was fair to substantial (Kg = 0.26 to 0.75) in the trained group and fair to almost perfect in the untrained group (Kg = 0.37 to 0.86). Despite the higher range of agreement values for the medial malleoli in the untrained group, the mean level of agreement was slightly higher in the trained group.

The fact that the greatest effect of training was seen for the ASIS may be indicative of a large variation in placement of the examiners thumbs in the untrained groups. The ASIS is a relatively large anatomical landmark and without specific instructions to standardise assessment, inconsistencies may have developed between examiners. In contrast, the reliability of palpation of the medial malleoli faired well across both groups regardless of prior training. This is most likely due to the fact that it is usually a more obvious bony landmark and asymmetry is more obvious on observation.

Assessment of the SILA in the trained group revealed low levels of agreement for both inter-examiner reliability (Kg = 0.04) and intra-examiner reliability (Kg = -0.04 to 0.64) and the trained group produced only slightly higher levels of agreement than the untrained group. The intra-examiner reliability produced a wide range of agreement possibly indicating error in landmark location among some examiners and varying degrees of palpatory skill or familiarity with palpation of the SILA. In addition, the SILA was the final landmark palpated on all participants and after lying still for a prolonged period of time they may have began making small adjustments to their position. This may have lead to changes in the symmetry of the landmarks thereby further reducing the agreement between examiners. Despite this, the levels of agreement concurred closely with the results of O'Haire's[9] study which found less than chance to substantial intra-examiner agreement (Kg = -0.05 to 0.69) and slight inter-examiner agreement (Kg = 0.08) for palpation of the SILA.

Interestingly intra-examiner reliability for the PSIS was slight to almost perfect, however inter-examiner reliability was only slight regardless of training, indicating that the examiners had an acceptable degree of consistency within themselves but not between each other. This trend is consistent with OHaire's[9] study which demonstrated slight to moderate intra-examiner agreement (Kg = 0.33) and slight inter-examiner agreement (Kg = 0.04) for palpation of the PSIS.

With growing evidence of the poor reliability of static palpation, the osteopathic profession needs to review the use of these practices despite the recommendation by many authors and the apparently frequent use by practicing osteopaths. Previous research suggests that pelvic asymmetry may actually have no association with low back pain, hence the validity of static palpation is highly questionable. Levangie *et al.*[14] investigated the association between low back pain of less than twelve months duration and pelvic asymmetry. They found that pelvic asymmetry was not positively associated with low back pain in any way that seemed clinically meaningful. Asymmetry of the PSIS revealed evidence of a weak positive association with low back pain, however they concluded that in the absence of a meaningful positive association between pelvic asymmetry and low back pain, evaluation and treatment strategies based on that premise should be reconsidered.

While this study demonstrated acceptable levels of intra-examiner reliability for the ASIS, PSIS and medial malleoli by the trained examiners, acceptable inter-examiner reliability of static palpation was not established and constitutes a challenge for future research. Future studies should consider examining only one landmark per session in an attempt to reduce any differences in agreement due to either examiner fatigue, or changes in the position of the participants. The training sessions did not appear to have a notable effect on the level of agreement, hence it is likely that only one training session would be sufficient in future studies. This study may have benefited from examining the effect of training on the same group of examiners. This may have been possible by conducting both a pre-training and post training study in order to observe any increases in agreement due to the training session alone, rather than variable levels of palpatory skill. It was assumed that both groups had equal palpatory ability before the study began, however this may not have been the case. As a result, it is possible that greater levels of agreement may have been achieved by using more experienced examiners.

CONCLUSION

The levels of intra-examiner reliability were higher than inter-examiner reliability, which concurs with previous research in this area. The training sessions produced a marginal overall increase in agreement both between and within examiners, however there were a small number of exceptions. The intra–examiner agreement was acceptable in the trained groups and ranged from slight to almost perfect, however inter-examiner agreement was generally poor whereby the highest level of agreement was fair for both the ASIS and medial malleoli. Palpatory examination currently plays an integral role in the diagnostic process within the manual medicine profession despite the growing evidence of poor reliability of static palpation of bony landmarks. Consequently further research should be aimed at devising methods to increase the levels of inter-examiner reliability, or the profession needs to reconsider the use of these examination methods.

References

1. Ward RC. Palpation and Clinical Problem Solving: A Discussion of the Palpatory Art with some Correlative Science. Australasian Musculoskeletal Medicine. 2000; 21-27.

2. Peace S, Fryer G. Methods used by members of the Australian osteopathic profession to assess the sacroiliac joint. Journal of Osteopathic Medicine. 2004; 7(1):25-32.

3. Greenman PE. Principles of Manual Medicine (2nd Ed). Baltimore: Williams and Wilkins; 1996:14,321-327.

4. DiGiovanna EL, Schiowitz S. An Osteopathic Approach to Diagnosis and Treatment. Pennsylvania: J.B. Lippincott Company; 1991:30,39-41.

5. Mitchell, FL and Mitchell, PKG. The Muscle Energy Manual Vol 1: Evaluation and Treatment of the Pelvis and Sacrum. Michigan, MET Press; 1999:63

6. Haas M. The Reliability of Reliability. Journal of Manipulative and Physiological Therapeutics. 1991;14: 199-208

7. Mior SA, McGregor DC, Schut B. The Role of experience in clinical accuracy. Journal of Manipulative and Physiological Therapeutics. 1990;13:68-71.

8. Gerwin RD, Shannon S, Hong CZ, Hubbarb D, Gevirtz R. Interrater reliability in myofascial trigger point examination. Pain. 1997;69:65-73

9. O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. Manual Therapy. 1999; 5:13-20.

10. Keating JC, Bergman TF, Jacobs GE, Finer BA, Larson K. Inter-examiner Reliability of Eight Evaluative Dimensions of Lumbar Segmental Abnormality. Journal of Manipulative and Physiological Therapeutics. 1990;13:463-470.

11. McConnell DG, Beal MC, Dinnar U and co-workers. Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. Journal of the American Osteopathic Association. 1980;79:59-68.

12. Haas M. Statistical Methodology for Reliability Studies. Journal of Manipulative and Physiological Therapeutics. 1991;14:119-132.

13. Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics.1997; 33:159-174.

14. Levangie PK. The association between static pelvic asymmetry and low back pain. Spine. 1999;24:1234-1242.

Table 1. Interpretation of Kappa values[13]

Value of Kappa	Agreement
0.0 - 0.2	Slight
0.21 - 0.4	Fair
0.41- 0.6	Moderate
0.61 – 0.8	Substantial
0.81 – 0.99	Almost perfect

Table 2. Intra-examiner reliability

		Mean Kappa	Kappa Range
ASIS	Trained	0.65	0.54 - 0.80
	Untrained	0.19	-0.01 – 0.40
Medial Malleoli	Trained	0.68	0.26 – 0.75
	Untrained	0.59	0.37 – 0.86
PSIS	Trained	0.54	0.12 – 0.83
	Untrained	0.49	0.22 – 0.76
SILA (Sup/Inf)	Trained	0.20	-0.10 – 0.64
	Untrained	0.03	-0.17 – 0.28
SILA (Ant/Post)	Trained	0.07	-0.04 – 0.39
	Untrained	0.20	-0.08 – 0.64

Table 3. Inter-examiner reliability

		Po	Pe	Kg	P
ASIS	Trained	0.58	0.45	0.24	<0.01
	Untrained	0.34	0.35	-0.01	<0.01
PSIS	Trained	0.53	0.49	0.08	<0.01
	Untrained	0.59	0.52	0.15	<0.01
Medial	Trained	0.54	0.34	0.31	<0.01
Malleoli	Untrained	0.53	0.35	0.28	<0.01
SILA	Trained	0.42	0.39	0.04	<0.01
(Sup/Inf)	Untrained	0.44	0.44	-0.01	<0.01
SILA	Trained	0.40	0.37	0.04	<0.01
(Ant/Post)	Untrained	0.38	0.38	-0.01	<0.01

Po = proportion of observed agreement, Pe = proportion of chance agreement, Kg = agreement beyond chance agreement, *P* = significance beyond chance agreement.