GIA98002

Title

The reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low back pain

Authors

Simon D. French B.App.Sc.(Chiro)

Sally Green B.App.Sc.(Physio), P.Grad.Dip.(Manip.Physio)

Andrew Forbes B.Sc (Hons), PhD

Research conducted as part of primary author's requirements to complete the degree of Master of Public Health at the Department of Epidemiology and Preventive Medicine, Monash University, Victoria, Australia.

Author for correspondence

Simon French

70 Wombat Drive

Eltham North VIC 3095

Australia

Ph: +61 3 9439 1798

simsime@ozemail.com.au

Source of support

Grant in Aid from the Australian Spinal Research Foundation

Running Title

Reliability of chiropractic diagnosis

Abstract

Objective

To assess the intra- and interexaminer reliability of a multidimensional spinal diagnostic method commonly utilised by chiropractors.

Design

An intra- and interexaminer Latin square, repeated measures reliability study. The techniques of diagnosis under investigation included visual postural analysis, pain description by the patient, plain static erect x-rays of the lumbar spine, leg length discrepancy, neurological tests, motion palpation, static palpation and orthopaedic tests.

Participants

Three experienced chiropractors examined 19 subjects, and 2 experienced chiropractors examined 10 and 9 subjects respectively, who were suffering from chronic mechanical low back pain.

Results

Intraexaminer reliability of the decision to manipulate a certain spinal segmental level was moderate (K = 0.47). The interexaminer agreement pooled across all spinal joints indicated fair agreement (K = 0.27). Interexaminer reliability for individual examiner pairs for the L4/L5 segmental level was slight (K = 0.09). At the L5/S1 level, the interexaminer reliability was fair (K = 0.25). For the sacroiliac joints, interexaminer reliability was slight (K = 0.04 and 0.14).

Conclusion

This study of commonly utilised chiropractic diagnostic methods in patients with chronic mechanical low back pain to detect manipulable lesions in the lower thoracic spine, lumbar spine and the sacroiliac joints, has revealed that the measures are not reproducible. The implementation

of these examination techniques alone should not be seen by practitioners to provide reliable information concerning where to direct a manipulative procedure in patients with chronic mechanical low back pain.

Key Indexing Terms

Observer variation, reproducibility of results, low back pain, lumbar vertebrae, chiropractic.

Introduction

Although the effectiveness of spinal manipulative therapy for low back pain has received some support in the literature (1, 2), the target of the manipulative procedure remains an elusive entity.

Chiropractors use a variety of techniques in the examination of a patient with musculoskeletal related pain to determine if and where spinal manipulation should be directed. Chiropractors claim they are able to detect which joints of the spine are in need of manipulation with the use of certain diagnostic criteria (3, 4). If clinical decisions are to be made on the basis of these findings, it is important that the examination procedures are both reliable and valid.

The reliability of a test is its consistency of results on repeated applications in a person with a given level of a disorder. Reliability depends on the variability of the disorder in the patient population of interest, and on the variability in the method of measurement and the skill with which it is made (5).

Intraexaminer reliability is the ability of one examiner to agree with him or herself with repeated measures on the same subject. Interexaminer reliability is the consistency of the results of a diagnostic test between one examiner and another, when the diagnostic test is performed on the same subject. All the sources of error in intraexaminer reliability are contained within interexaminer reliability, in addition to any differences that may arise between examiners. Therefore, to demonstrate that a test is adequate with respect to reliability, the demonstration of high interexaminer reliability is all that is required. If interexaminer reliability is not high, then intraexaminer reliability should be explored to assist in the location of the source of unreliability (6).

A 1992 survey was designed to determine which techniques chiropractors commonly utilised to ascertain the presence of manipulable lesions (7). These techniques are listed as follows:

1. Visual postural analysis
2. Pain description of the patient.
3. Plain static erect x-rays of the lumbar spine.
4. Leg length discrepancy.

5. Neurological tests.
6. Motion palpation.
7. Static palpation.
8. Orthopaedic tests.

Chiropractors in Victoria frequently used the above techniques to determine the presence of manipulable lesions and believed that the techniques were reliable. They also commonly used these procedures in combination (8).

Some of these procedures have been examined extensively for reliability, including x-ray analysis (9-14), leg length check (15-20), motion palpation (21-29) and static palpation (30, 31). To date, the reliability of each of the above examination procedures used in isolation has not been established. For each of the procedures, further research is justified with well-designed studies utilising patients who are representative of those seen in chiropractic practices.

The assessment of the reliability of the decision to manipulate a particular vertebral segment in patients with chronic low back pain is lacking. This decision has been inferred by some of the studies but the clinical decision itself has only been assessed in 2 studies. McConnell et al. (32) utilised only descriptive statistics, and therefore limited conclusions can be drawn from this study. Thus, only one published study has been adequately designed to assess the reliability of the decision to manipulate a vertebral segment, and this study concluded poor interexaminer reliability (28).

The current study was employed to determine the reliability of a group of examination procedures that are commonly utilised by chiropractors, that are commonly utilised in combination and that the chiropractic profession believe are reliable. The study was undertaken in a group of subjects suffering from chronic mechanical low back pain.

Materials and Methods

The method of this study was approved by the Monash University Standing Committee on Ethics in Research on Humans. The purpose of the study was explained to each of the participants

by way of a plain language statement. Patients and chiropractors were then required to complete an informed consent form prior to he or she agreeing to be involved in the study.

Subjects

A total of 5 chiropractors who commonly employ the methods listed above to determine manipulable lesions, and are registered in the state of Victoria, Australia, were chosen as the subset of examiners. Four chiropractors were used on the first day of the study, and 4 on the second day of the study. Three chiropractors were involved for both days, and 2 chiropractors only completed the study on one of the days.

The subjects included in this study were a group of 19 males and females suffering from chronic mechanical low back pain. Chronic mechanical low back pain is defined as pain within the boundaries of the lower costal margins superiorly, the gluteal folds inferiorly, and at least 3 months in duration (33). It was required that the patients were older than 18 years old, and that their condition was non-irritable (34) to minimise the chance that the examination procedure would aggravate symptoms. Exclusion criteria included the following: acute exacerbation of the chronic complaint on the days leading up to the study; refusal to gain informed consent; radiography indicating more or less than 5 lumbar vertebrae or missing spinal processes; and any contraindications to spinal manipulation.

Chronic low back pain is a common characteristic of the patients who visit and receive manipulation from chiropractors (35). Thus the use of subjects with chronic low back pain ensures the utilisation of representative chiropractic patients.

Sample size for reliability studies involves decisions concerning both the number of patients and the number of examiners. With 4 chiropractors and 20 patients, the study has approximately 80% power to detect kappa statistics of approximately 0.40 using a 5% significance level (36).

Method

Each patient completed a series of questionnaires to determine their demographics and the characteristics of their low back pain. These questionnaires included a visual analogue scale and the Oswestry back pain questionnaire (37).

The chiropractic examiners were blind to the purpose of the investigation to minimise bias in relation to the evaluation of intraexaminer reliability. The chiropractors were told that the purpose of the study was to investigate diagnostic procedures performed by chiropractors, but not specifically that the study was a reliability study.

The study was conducted over a 2 day period. The first day each of the 4 chiropractors examined each of the 10 patients on 2 occasions. On the second day each of the 4 chiropractors examined another 9 patients on 2 occasions. Each patient was required for one half day, minimising any inconvenience to the patients, and ensuring that each patient was examined in a stable state of symptomatology.

A Latin square design accounts for the number of subjects and examiners, and the order of the testing (38). The Latin square design allows for any order effect that may occur, such as change of pain or mobility produced by repeated testing. A replicated 4 by 4 Latin square design was developed, with each block of 4 patients making up a square. Each subject was examined by each examiner, with examiners evenly distributed across the order of examination, so that, for example, the number of patients for which examiner A was the first examiner was the same as the number for examiner B. A table with 10 subjects and 4 examiners was used the first week, and a table with 9 subjects and 4 examiners was used the second week. The subjects were randomly assigned a number and the examiners were randomly assigned a letter, which directed where subjects were assigned within the table, and in what order they were to be examined.

The examination procedure was conducted in closed rooms to ensure that the examiners were blind to each other's decisions. For each of the joints of the lower thoracic spine, lumbar spine, and the sacroiliac joints, the examiners were required to indicate the likelihood that a particular joint did or did not require manipulation according to the examination procedure they had just performed. Each examiner examined each of the subjects once, and then examined each of the subjects again later in the day. The order of examinations remained the same as the initial randomisation procedure to account for any differences in relation to fatigue of the examiner or the subject, and also to keep consistent any changes that may have occurred in the subject's spine as the examination procedures were carried out.

The following examination procedures were carried out:

1. **Visual postural analysis**. This was performed with the patient in the standing position, to determine any postural findings that the examiner believed would assist them in the detection of the manipulable lesion/s.

2. **Pain description of the patient**. The chief investigator (SDF) obtained this information from each subject before the day of data collection. This included a history of each subject's low back pain, including the site of pain, the nature of the pain, the length of time the pain has been experienced, and aggravating and relieving factors. This summary was available for each of the examiners. The examiners were able to ask the subject any further questions they felt were necessary in order to accurately detect any manipulable lesions.

3. **Plain static erect x-rays of the lumbar spine**. If available, plain static x-rays were presented for the chiropractors to view. No subject was x-rayed for the purposes of this study. If they deemed it was necessary, the examiner viewed the x-rays in order for this information to assist them in determining which joints of the spine require manipulation.

4. **Leg length discrepancy**. This was measured by the examiner's preferred technique, either supine, prone or both to assist them in deciding the level and side of any manipulable lesions.

5. **Neurological tests**. If the examiner decided that neurological examination, including assessment of myotomes, dermatomes and reflexes, was required, they were able to perform any neurological test at their discretion in order to assist them in determining the level and side of any manipulable lesions.

6. **Motion palpation**. It was up to the discretion of the examiner to choose which technique/s, either passive or active motion palpation, they preferred in order to assist them to detect any manipulable lesions.

7. **Static palpation**. This was used to determine any tenderness, muscle hypertonicity, misalignment and any other palpatory finding that assisted the examiner to determine the level and side of the manipulable lesion/s.

8. **Orthopaedic tests (including range of motion).** The examiners were able to perform any orthopaedic tests (for example straight leg raise, Lewin Gaenslen's test) they deemed suitable in order to determine the level and side of the manipulable lesion/s.

Analysis

The primary outcome investigated was the final decision of the examiner of where manipulation was to be directed. Following completion of the above tests, the examiner was to decide yes or no for each of the joints of the lower thoracic and lumbar spine and the sacroiliac joints. For each segmental level of the spine, an examiner could choose to manipulate either the right or left side, both sides or neither side.

A number of analyses were performed. Firstly, overall intraexaminer and interexaminer reliability was calculated. This was determined by calculating kappa values for each examiner and each examiner pair, pooling the examiners' decisions across all the spinal joints examined. Separate reliability scores for each of the joints of the lower thoracic and lumbar spine and the sacroiliac joints were also calculated.

In addition, for interexaminer reliability, the spinal segmental levels of the lower lumbar spine were collapsed into regions. Two adjacent segmental levels were collapsed to account for the possibility that examiners may have identified the same manipulable lesions, but they recognised it as a different spinal level.

Generalised kappa can be used to evaluate reliability in this case where more than 2 examiners are categorising patients, to determine the agreement that occurs beyond that expected by chance alone (39). Generalised kappa is the weighted average of pairwise kappas, with lower weights given to pairs of raters where the expected agreement by chance is high. In addition, kappa scores were calculated for each combination of examiner pairs. All analyses were performed using version 5.0 of the Stata statistical software package (40).

The guidelines of Landis and Koch (41) were utilised for interpreting the kappa values obtained (Table 1).

Under certain conditions kappa becomes unstable as a reliability statistic. This problem is called limited variation and occurs when there is a large proportion of agreement, but most of the

agreement is limited to only one of the possible rating choices. For example, if the prevalence of manipulable lesions is very low in a spinal segmental joint, then the examiners will agree with each other that there is no manipulable lesion. However, if one examiner decides that there is a manipulable lesion present, this one decision under limited variation can make the difference between a study demonstrating poor and excellent reliability (39). In recognition of this limitation, percentage agreement values are presented in addition to the kappa values.

Results

Examiner's Details

Each of the chiropractors obtained their chiropractic qualifications of a Bachelor of Applied Science (Chiropractic) from the same institution, Phillip Institute of Technology, now known as Royal Melbourne Institute of Technology. One of the chiropractors also had nursing qualifications and 1 had post-graduate qualifications. Examiner A had 12 years experience in clinical practice, Examiner B 5 years, Examiner C 17 years, Examiner D 18 years and Examiner E 5 years (mean 11.4 years).

Two of the chiropractors (Examiners A and D) examined all 19 patients, 1 chiropractor (Examiner B) examined 18 patients, 1 chiropractor (Examiner C) examined 10 patients and 1 chiropractor (Examiner E) examined 9 patients.

The attempt to blind the examiners to the purpose of the study was not successful. At the conclusion of the second day of the study, 4 of the 5 chiropractors stated that they had assumed that the study was a reliability study.

Patient Details

A total of 19 subjects were enrolled into the study. Ten subjects attended for the study on the first week and 9 the second week. Two subjects withdrew from the study on the day of the second week, and only 1 other subject was available at short notice. On the second day of the study, 1 subject was examined by only 3 of the 4 chiropractors in the second round. The subject chose to leave early because his low back pain was increasing with each examination. Table 2 summarises the demographic data of the subjects enrolled into the study.

Prevalence of Manipulable Lesions

Table 3 outlines the frequency that each joint was chosen for manipulation, and is the prevalence of manipulable lesions in each joint examined. The low prevalence of manipulable lesions in the lower thoracic and upper lumbar spine will prohibit the usefulness of the kappa

statistics calculated in the following sections, because limited variation is present at these levels. Interpretation of the results regarding this region of the spine should thus be treated with caution.

Intraexaminer Reliability

Table 4 illustrates the percentage agreement and the kappa values calculated for overall intraexaminer reliability of each of the examiners pooled across all the spinal joints examined. For the decision of which segmental level to manipulate, examiner A achieved the highest kappa score, with substantial agreement (K = 0.73). Examiner C had the lowest kappa score with only slight agreement (K = 0.13). All intraexaminer kappa scores were statistically significant except for Examiner C. The mean kappa score for all the examiners was 0.47, indicating moderate intraexaminer reliability.

The results in Table 5 are the intraexaminer reliability data presented by joint. The mean kappa scores for each joint indicate that, in the context of the examination protocol utilised, the left sacroiliac joint has the highest intraexaminer reliability score, with moderate agreement being achieved within the examiners (K = 0.43, range -0.18 to 0.75). Poor intraexaminer agreement was found at T11/T12 (K = -0.08, range -0.06 to -0.09). Slight intraexaminer agreement was found at T12/L1 (K = 0.07, range -0.11 to 0.31), L1/L2 (K = 0.00, range 0.0 to 0.0) and L3/L4 (K = 0.16, range 0.00 to 0.64). Fair intraexaminer agreement was found at L2/L3 (K = 0.23, range -0.09 to 1.0), L4/L5 (K = 0.23, range -0.11 to 0.86), L5/S1 (K = 0.26, range -0.21 to 0.56) and the right sacroiliac joint (K = 0.38, range -0.18 to 0.75). The results for the thoracolumbar spine region (T11/T12 to L3/L4) must be interpreted carefully considering the high percentage agreement but low kappa.

Interexaminer Reliability

Table 6 gives the mean percentage agreement and generalised kappa values for all the joints of the thoracic and lumbar spine and the sacroiliac joints, together with the range of the pairwise kappa values formed from all pairs of examiners.

Segmental reliability

The generalised kappa score for all the examiner pairs indicated fair interexaminer agreement ($K = 0.27$, $p < 0.0001$), with the range of pairwise kappa values indicating slight to fair agreement ($K = 0.18$ to 0.36).

The data was broken down by segmental level to determine the interexaminer reliability at the individual spinal segmental levels assessed by the chiropractors. The data for the segmental levels of the lower thoracic and upper lumbar spine (T11/T12 through to L3/L4) are not included due to limited variation occurring. This data is not discussed further because it is considered that the measurement of reliability at these segmental levels is untenable due to the low prevalence of manipulable lesions. This is indicated by the high percentage agreement but low Kappa scores due to most examiners agreeing that there is not a manipulative lesion present.

The other segmental levels of lumbar spine and the sacroiliac joints appear to have relatively stable Kappa statistics and the results are presented in Table 6.

For the decision to manipulate at the L4/L5 segmental level, the generalised kappa score indicated slight interexaminer reliability ($K = 0.09$, $p = 0.08$) and this was not statistically significant. The range of pairwise kappa values indicated poor to fair agreement (range -0.07 to 0.40).

For the decision to manipulate at the L5/S1 level, the generalised kappa value for L5/S1 indicated fair interexaminer reliability ($K = 0.25$, $p < 0.0001$), and the range of pairwise kappa values indicated slight to moderate agreement (range 0.0 to 0.47).

The generalised Kappa score for the decision to manipulate the right sacroiliac joint indicated slight reliability ($K = 0.04$, $p = 0.30$) and was not statistically significant. The range of pairwise kappa values indicated poor to fair agreement (range -0.20 to 0.36). The generalised Kappa score for the decision to manipulate the left sacroiliac joint also indicated slight interexaminer reliability ($K = 0.14$, $p = 0.02$), and the range of pairwise kappa values indicated poor to moderate agreement (range -0.37 to 0.57).

Joint Reliability

The data was then broken down to the decision of whether to manipulate a specific spinal joint, i.e. which side of the spinal segment requires manipulation.

The results for the interexaminer reliability for examination of specific joints of the lumbar spine were also disappointing. The generalised Kappa scores also indicated poor to slight interexaminer reliability (range K = -0.16 to K = 0.08), and none of these kappa values were statistically significant. The range of pairwise kappa scores also indicated poor to slight interexaminer reliability (range K = -0.04 to K = 0.11).

Collapsed segmental levels

Two adjacent spinal segmental levels were collapsed to account for the examiners differing in their labelling of a particular spinal level. The results are presented in Table 6. For the L3-L4-L5 collapsed data, the generalised kappa score indicated slight interexaminer reliability (K = 0.06, p = 0.21), and the range of pairwise kappa scores indicated poor to fair agreement (range K = -0.10 to 0.33). For the L4-L5-S1 collapsed data, the generalised kappa score also indicated slight interexaminer reliability (K = 0.08, p = 0.38), and the range for pairwise kappa values indicated poor to fair agreement (range K = -0.16 to 0.29). The kappa values for the collapsed data were not statistically significant.

Discussion

This study of commonly utilised chiropractic diagnostic methods to detect manipulable lesions in the thoracic and lumbar spine and the sacroiliac joints in patients with chronic mechanical low back pain has revealed that the measures are not reproducible. The decision to manipulate was not reproducible either by the same examiners on different occasions, or by different examiners on the same occasion. Based on the results of this study, the use of these examination techniques in combination to detect manipulable lesions should not be seen by practitioners to provide reliable information concerning where to direct a manipulative procedure.

The mean intraexaminer agreement for concordance on the level to manipulate was moderate. The intraexaminer agreement for concordance on the level to manipulate for individual examiners ranged from slight to substantial. With the data reduced to the individual spinal segmental level, moderate intraexaminer agreement was achieved at the left sacroiliac joint, fair agreement at L2/L3, L4/L5, L5/S1 and the right sacroiliac joint, slight agreement at T12/L1, L1/L2 and L3/L4 and poor agreement at T11/T12.

The overall interexaminer reliability for the entire lower thoracic and lumbar spine, and the sacroiliac joints, was fair. The reliability of the individual spinal segments of the lower thoracic and upper lumbar spine was not assessed due to limited variation. For the lower lumbar spine, the reliability of the L4/L5 level was slight, and the L5/S1 level fair. For the individual joints of the lower lumbar spine, the interexaminer reliability was poor to slight. Collapsing of the data to evaluate the reliability over a region of 2 joints did not improve the agreement between examiners. The reliability of the both the right and left sacroiliac joints was slight.

A paradox of high percentage agreement but low kappa value can occur because kappa coefficients represent the proportion of agreement after chance agreement is excluded (42). If there is a disproportionately large number of agreements of absent manipulable lesion, many of the agreements of absent can be expected by chance. Thus, an extremely high degree of agreement between examiners is required in order to obtain a high kappa value. This is the situation that occurred in the lower thoracic and upper lumbar spine data in this study. The results for this

region, and also overall results that include this region in the calculation, must be interpreted with caution.

There is difficulty in controlling for conscious and unconscious memory cues that make it virtually impossible to ensure sufficient blinding of examiners when assessing intraexaminer reliability (43). Intraexaminer reliability may be affected by the chiropractors remembering their decision in the previous examination of a particular patient. Steps were undertaken to minimise this. Firstly, for each spinal level, a decision was to be made about whether that segmental level requires manipulation, and which side of the segmental level requires manipulation. A total of 14 joints are present in this region, and the examiner was required to make a number of decisions about each of them. Each segmental level in the spine had the possible combinations of no-no, no-yes, yes-no or yes-yes. This large number of decisions made it less likely that chiropractors were able to remember their earlier decision made for each of the 19 patients when they examined them for a second time.

Secondly, each examiner was blinded to the purpose of the study. Unfortunately this blinding was not effective as 4 of the 5 examiners had assumed that the study was in fact a reliability study. The fact that the intraexaminer reliability was greater than the interexaminer reliability makes it difficult to exclude memory bias as a cause of this result.

The reliability of clinicians to identify spinal levels in the lumbar spine is not established (44). Not reliably identifying a spinal level between the different examiners may be the sole cause for disagreement of where to manipulate the spine. However, the collapsing of the data to regional levels accounted for this potential variability as the cause of disagreement in the lower lumbar spine. No improvement in reliability coefficients strengthens the conclusions of this study.

Sackett et al (45) outline possible aetiologies of clinical disagreement. Firstly, clinical disagreement may arise from the examiner. This reliability study relied upon the examiners' senses of sight, hearing and touch. All examiners had adequate sight and hearing, although it may be possible that a different ability to palpate and feel for manipulable lesions is one reason why the diagnostic regimen was not reliable. There may also be a tendency to record inference rather than evidence. Examiners may have picked up certain cues from the patient in the early part of the

examination and decided upon the presence of manipulable lesions in certain segments of the spine, before the appropriate examination had been carried out. The third aetiology outlined by Sackett in relation to the examiner is ensnarement by diagnostic classification schemes. The diagnostic criteria of a manipulable lesion has not been outlined and specified in the literature to date. This may lead to arbitrary classification by individual examiners of what constitutes the presence of a manipulable lesion in a spinal joint. Future studies may consider gaining a consensus view prior to the study's commencement of what findings in the examination procedure would constitute the presence of a manipulable lesion.

The next aetiology is entrapment by prior expectation. The presence of manipulable lesions in the spine of a patient with low back pain would be expected by a chiropractor. Examiners have expectations that, firstly, a manipulable lesion will be present if the pain is considered to be mechanical from the patient's history, and secondly that manipulable lesions are more commonly found in certain areas of the spine. The detection of manipulable lesions is based upon a series of subjective examinations, and this cause of clinical disagreement is thus a high possibility. The final aetiology is simple incompetency. All the examiners had ample experience and qualifications in the detection of manipulable lesions, and therefore this aetiology is not expected.

The clinical disagreement may also arise from the subject being examined. Biologic variation in the system being examined is a possible aetiology of clinical disagreement in this study. A subject having their low back examined a number of times in one day may lead to a change in the function of their spine, and may lead to a variation of the presence of manipulable lesions in the spine. However, the biologic plausibility of this explanation must be questioned. If the manipulable lesion is such that its presence can be influenced by interventions other than manipulation, then perhaps the definition itself needs further clarification.

The examination itself may also be a cause of examiner disagreement. In order to make this study clinically relevant to practising chiropractors, no training prior to the implementation of the examination procedure was carried out. The procedure for examination in this study was not practiced or agreed upon by the examiners. This may have led to the examiners interpreting the protocol in different ways, and thus performing procedures differently to their colleagues.

However, this is what occurs in clinical practice, and the study was undertaken for precisely this reason. Future studies may consider training of the examiners prior to the study's implementation to establish consensus views on the examination procedure.

Is it clinically important that chiropractors detect a specific joint for manipulation for the procedure to be successful? Traditional chiropractic theory suggests that this is the case (46). Two issues on the topic of validity are relevant for discussion here. The first is in regards to the validity of the detection of manipulable lesions. No study has been conducted to evaluate the validity of the presence of manipulable lesions in the lumbar spine. Manipulable lesions may be a figment of the collective chiropractic, and other physical therapy professions', imagination. Reliability of such a lesion will be proven again and again to be inadequate. The second is the form of treatment that is employed when manipulable lesions are detected. Chiropractors claim to be able to manipulate specific joints in the spine. The controlled clinical trials that have been performed on manipulation for chronic low back pain have not evaluated this aspect of the procedure (47-51). Further research is required in this form of therapy for low back pain to evaluate if it is clinically important to determine if 'specific' manipulation to manipulable lesions is more or less effective than 'non-specific' manipulation.

Conclusion

The assessment of commonly utilised chiropractic diagnostic methods to detect manipulable lesions in the lower thoracic and lumbar spine, and the sacroiliac joints, in subjects with chronic mechanical low back pain, has revealed that the measures are not reproducible. The decision to manipulate was not reproducible either by the same examiners on different occasions, or by different examiners on the same occasion. Based on the results of this study, the use of these examination techniques to detect manipulable lesions in subjects with chronic mechanical low back pain should not be seen by practitioners to provide reliable information concerning where to direct a manipulative procedure.

Acknowledgments

We thank the Australian Spinal Research Foundation for their financial support of this study.

References

1. Koes BW, Assendelft WJJ, van der Heijden GJMG, Bouter LM. Spinal manipulation for low back pain: an updated systematic review of randomised clinical trials. Spine 1996;21:2860-73
2. Shekelle PG, Adams AH, Chassin MR, Hurwitz EL, Brook RH. Spinal manipulation for low back pain. Ann Int Med 1992;117:590-8
3. Dishman RW. Static and dynamic components of the chiropractic subluxation complex. J Manipulative Physiol Ther 1988;11:98-107
4. Faye LJ, Wiles MR. Manual examination of the spine. In: Haldeman S, ed. Principles and practice of chiropractic. Connecticut: Appleton and Lange, 1992:301-18.
5. Morrison AS. Screening. In: Rothman KJ, Greenland S, eds. Modern epidemiology. Philadelphia: Lippincott-Raven Publishers, 1998:499-517.
6. Streiner DL, Norman GR. Health measurement scales: a practical guide to their use and development. New York: Oxford University Press Inc., 1995.
7. Walker BF, Buchbinder R. Most commonly used methods of detecting subluxation and the preferred term for its description: a survey of chiropractors in Victoria, Australia. J Manipulative Physio Ther 1997;20:583-8
8. Walker BF. Most common methods used in combination to detect spinal subluxation: a survey of chiropractors in Victoria. Aust Chiro Osteo 1998;7:109-11
9. Plaugher G, Hendricks AH. The inter- and intraexaminer reliability of the Gonstead pelvic marking system. J Manipulative Physiol Ther 1991;14:503-8
10. Reinert OC. An analytical survey of structural aberrations observed in static radiographic examinations among acute low back cases. J Manipulative Physiol Ther 1988;11:24-30
11. Phillips RB, Frymoyer JW, McPherson BV, Newburg AH. Low back pain: a radiographic enigma. J Manipulative Physiol Ther 1986;9:183-7

12. Assendelft WJJ, Bouter LM, Knipschild PG, Wilmink JT. Reliability of lumbar spine radiograph reading by chiropractors. Spine 1997;22:1235-41
13. Rhudy TR, Sandefur MR, Burk JM. Interexaminer/intertechnique reliability in spinal subluxation assessment: a multifactorial approach. Am J Chiro Med 1988;1:111-4
14. Haas M, Nyiendo J, Peterson C, et al. Interrater reliability of roentgenological evaluation of the lumbar spine in lateral bending. J Manipulative Physiol Ther 1990;13:179-89
15. Venn EK, Wakefield KA, Thompson PR. A comparative study of leg-length checks. Eur J Chiro 1983;31:68-80
16. Shambaugh JP, Sclafani L, Fanselow D. Reliability of the Derifield-Thompson test for leg length inequality, and use of the test to demonstrate cervical adjusting efficacy. J Manipulative Physiol Ther 1988;11:396-9
17. Fuhr AW, Osterbauer PJ. Interexaminer reliability of relative leg length evaluations in the prone, extended position. Chiro Tech 1989;1:13-8
18. Youngquist M, Furh AW, Osterbauer PJ. Interexaminer reliability of an isolation test for the identification of upper cervical subluxation. J Manipulative Physiol Ther 1989;12:93-7
19. Falltrick DR, Pierson SD. Precise measurement of functional leg length inequality and changes due to cervical spine rotation in pain free students. J Manipulative Physiol Ther 1989;12:364-8
20. Rhodes DW, Mansfield ER, Bishop PA, Smith JF. Comparison of leg length inequality measurement methods as estimators of the femur head height difference on standing x-ray. J Manipulative Physiol Ther 1995;18:448-52
21. Gonnella C, Paris SV, Kutner M. Reliability in evaluating passive intervertebral motion. Phys Ther 1982;62:436-44
22. Bergstrom E, Courtis G. An inter- and intra-examiner reliability study of motion palpation of the lumbar spine in lateral flexion in the seated position. Euro J Chiropractic 1986;34:121-41
23. Jull G, Bullock M. A motion profile of the lumbar spine in an ageing population assessed by manual examination. Physiotherapy Practice 1987;3:70-81

24. Love RM, Brodeur RR. Inter- and intraexaminer reliability of motion palpation for the thoracolumbar spine. J Manipulative Physiol Ther 1987;10:1-4

25. Boline PD, Keating JC, Brist J, Denver G. Interexaminer reliability of palpatory evaluation of the lumbar spine. Am J Chiro Med 1988;1:5-11

26. Mootz RD, Keating JC, Kontz HP, Milus TB, Jacobs GE. Intra- and interobserver reliability of passive motion palpation of the lumbar spine. J Manipulative Physiol Ther 1989;12:440-5

27. Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. Phys Ther 1994;74:801-11

28. Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. Phys Ther 1995;75:786-95

29. Herzog W, Read LJ, Conway PJW, Shaw LD, McEwen MC. Reliability of motion palpation procedures to detect sacroiliac joint fixations. J Manipulative Physiol Ther 1989;12:86-92

30. Keating JC, Bergmann TF, Jacobs GE, Finer BA, Larson K. Interexaminer reliability of eight evaluative dimensions of segmental abnormality. J Manipulative Physiol Ther 1990;13:463-70

31. Boline PD, Haas M, Meyer JJ, Kassak K, Nelson C, Keating JC. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality: Part II. J Manipulative Physiol Ther 1993;16:363-74

32. McConnell DG, Beal MC, Dinnar U, et al. Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. JAOA 1980;79:441-50

33. Frank A. Low back pain. BMJ 1993;306:901-9

34. Corrigan B, Maitland GD. Practical orthopaedic medicine. London: Butterworth & Co (Publishers) Ltd, 1988.

35. Nyiendo J, Phillips RB, Meeker WC, Konsler G, Jansen R, Menon M. Comparison of patients and patient complaints at six chiropractic college teaching clinics. J Manipulative Physiol Ther 1989;12:79-85

36. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Statist Med 1998;17:101-10

37. Fairbank JCT, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271-3

38. Armitage P, Berry G. Statistical methods in medical research. Cambridge: Blackwell scientific publications, 1994.

39. Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991;14:119-32

40. Stata Corporation. Intercooled stata for Windows 95. College Station, 1997.

41. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74

42. Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9

43. Haas M. The reliability of reliability. J Manipulative Physiol Ther 1991;14:199-208

44. McKenzie AM, Taylor NF. Can physiotherapists locate lumbar spinal levels by palpation? Physiotherapy 1997;83:235-9

45. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston: Little, Brown and Company, 1991.

46. Dishman RW. Static and dynamic components of the chiropractic subluxation complex. J Manipulative Physiol Ther 1988;11:98-107

47. Bronfort G, Goldsmith CH, Nelson CF, Boline PD, Anderson AV. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomised, observer blinded clinical trial. J Manipulative Physiol Ther 1996;19:570-82

48. Meade TW, Dyer S, Browne W, Townsend J, Frank AO. Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment. BMJ 1990;300:1431-7

49. Waagen GN, Haldeman S, Cook G, Lopez D, DeBoer KF. Short term trial of chiropractic adjustments for the relief of chronic low back pain. Manual Medicine 1986;2:63-7

50. Triano JJ, McGregor M, Hondras MA, Brennan PC. Manipulative therapy versus education programs in chronic low back pain. Spine 1995;20:948-55

51. Koes BW, Bouter LM, van Mameren H, et al. The effectiveness of manual therapy, physiotherapy, and treatment by the general practitioner for nonspecific back and neck complaints: a randomised clinical trial. Spine 1992;17:28-35

Table 1. *Guidelines of Landis and Koch for interpreting Kappa values* (41)

Kappa	Interpretation
< 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. *Patient characteristics.*

	Median	Mean	Std. dev	Range
Age (years)	41.5	41.3	11.2	24.6 – 67.7
Duration of low back pain (years)	14.0	16.2	9.8	2 – 35
VAS (max = 10)	5.0	4.4	2.2	1 – 8
Oswestry (max = 100)	36.0	35.0	13.2	14 – 58
Gender	M 14	73.7 %		
	F 5	26.3 %		

Table 3. *Prevalence of manipulable lesions* [a]

Segmental level	Frequency	Joint	Frequency
T11/12	6 (4.0)	T11/12 right	5 (3.3)
		T11/12 left	1 (0.7)
T12/L1 [b]	10 (6.6)	T12/L1 right	6 (4.0)
		T12/L1 left	3 (2.0)
L1/L2	2 (1.3)	L1/L2 right	2 (1.3)
		L1/L2 left	0 (0.0)
L2/L3	10 (6.6)	L2/L3 right	5 (3.3)
		L2/L3 left	7 (4.6)
L3/L4	9 (6.0)	L3/L4 right	6 (4.0)
		L3/L4 left	7 (4.6)
L4/L5	21 (13.9)	L4/L5 right	15 (10.0)
		L4/L5 left	15 (10.0)
L5/S1	86 (57.0)	L5/S1 right	65 (43.3)
		L5/S1 left	52 (34.7)
Right SI	61 (40.4)		
Left SI	43 (28.5)		

[a] Frequency and percentage (in parentheses) of manipulable lesion for spinal level. Examiners could choose right side of the joint, left side or both.

[b] One examiner indicated that they would manipulate the level, but did not indicate which side they would manipulate

Table 4. *Overall intra-examiner reliability*

Examiner	P_o	Kappa	p-value
A	92	0.73	$p < 0.0001$
B	82	0.38	$p < 0.0001$
D	87	0.59	$p < 0.0001$
C[a]	73	0.13	$p = 0.11$
E[b]	88	0.51	$p < 0.0001$
Mean		0.47	

P_o : Observed proportion of agreement (%)

[a] Examiner C assessed 10 patients

[b] Examiner E assessed 9 patients

Table 5. *Intra-examiner reliability for individual spinal segments* [a]

	Kappa (P_o)			Kappa (P_o)		
Joint	Examiner A	Examiner B	Examiner D	Examiner C	Examiner E	Mean
T11/T12	[b]	-0.09 (79)	-0.06 (89)	[b]	[b]	-0.08 (84)
T12/L1	[b]	0.0 (84)	0.31 (84)	-0.11 (80)	[b]	0.07 (83)
L1/L2	[b]	[b]	0.0 (95)	0.0 (90)	[b]	0.0 (92)
L2/L3	1.0 (100)	-0.09 (74)	[b]	0.0 (80)	0.0 (89)	0.23 (86)
L3/L4	0.64 (95)	[b]	0.0 (89)	0.0 (70)	0.0 (89)	0.16 (86)
L4/L5	0.86 (95)	-0.08 (84)	0.0 (95)	-0.11 (80)	0.5 (78)	0.23 (86)
L5/S1	0.56 (84)	0.11 (56)	0.45 (74)	-0.21 (30)	0.4 (78)	0.26 (64)
Right SI	0.16 (74)	0.43 (72)	0.43 (74)	0.4 (70)	0.5 (78)	0.38 (73)
Left SI	0.58 (84)	0.75 (89)	0.52 (79)	-0.18 (60)	0.5 (78)	0.43 (78)

[a] Kappa and percentage agreement (in parentheses) scores for each spinal segment examined

[b] Kappa scores not calculated due to very low prevalence of manipulable lesions

P_o : Observed proportion of agreement (%)

Table 6. *Interexaminer reliability* [a]

Joint	P_o		Kappa		
	Mean	Range	Generalised	p-value	Range [b]
Overall	78	75 to 81	0.27	$p < 0.0001$	0.18 to 0.36
L4/L5	77	65 to 89	0.09	$p = 0.08$	-0.07 to 0.4
L4/L5 R	83	68 to 95	0.03	$p = 0.31$	-0.1 to 0.44
L4/L5 L	78	63 to 95	0.08	$p = 0.11$	-0.1 to 0.34
L5/S1	64	22 to 76	0.25	$p < 0.0001$	0.0 to 0.47
L5/S1 R	51	17 to 74	0.01	$p = 0.45$	-0.25 to 0.3
L5/S1 L	48	17 to 82	-0.16	$p = 0.99$	-0.15 to 0.27
R SI	55	45 to 67	0.04	$p = 0.30$	-0.2 to 0.36
L SI	64	35 to 83	0.14	$p = 0.02$	-0.37 to 0.57
L3-L4-L5	68	45 to 84	0.06	$p = 0.21$	-0.10 to 0.33
L4-L5-S1	57	45 to 65	0.02	$p = 0.38$	-0.16 to 0.29

[a] Mean percentage agreement and Kappa values for individual segments and the sacroiliac joints

[b] The range is taken from the means of the pairwise kappa values