INTEREXAMINER RELIABILITY OF CHIROPRACTIC EVALUATION FOR CERVICAL SPINE PROBLEMS - A Pilot Study

Part 1: Graduates From One Institution

PETER J TUCHIN  BSc Grad Dip Chiro *
CHRISTOPHER J HART  MChiro †
CATRIONA JOHNSON  BSc †
ROBERT COLMAN  MChiro †
ADAM GEE  MChiro †
IAN EDWARDS  MChiro †
MARK PLUCKNETT  BSc †
ROD BONELLO  BSc DC DO MHA †

Abstract:
Objective: A pilot study was conducted to determine whether untrained examiners could agree on palpatory findings in the cervical spine.
Design: Fifty-three university students, (most of whom were chiropractic students), had their cervical spines examined by seven different chiropractors using their own clinical methods, of which motion palpation was a common, but not standard component.
Setting: Chiropractic Centre in Macquarie University.
Participants: Volunteer university students.
Main Outcome Measures: Individual clinical methods, which included static and/or motion palpation, vertebral springing, range of motion and applied kinesiology.
Results: Statistically, for the total group, there was poor interexaminer reliability. Of eight examiners, four did not disagree significantly, the next two examiners disagreed with each other but only at a single level and the remaining two examiners disagreed with most of the other examiners and each other.
Conclusion: In the cervical spine, it appears that C6 is the level of highest contention, followed by C1 and C5. Essentially the results suggest that combinations of examiners show reasonable consistency at identifying the same entity while using their own typical examination techniques. The nature of these palpable findings, leading to a diagnosis of subluxation or vertebral dysfunction is ill defined. Several issues were considered as important: expectations of examiners, research design, subject compliance, role of asymptomatic subjects and what the examiners were actually detecting.

Key Indexing Terms: Chiropractic, cervical spine, reliability, neck pain, manual medicine, musculoskeletal diagnosis.

INTRODUCTION

Back pain has been described as the "nemesis of medicine and the albatross of industry" (1). Increasingly, manual therapy is claiming to have substantial effects in the treatment of certain spinal conditions (2). To date, there is no substantiated method existing that determines whether a patient has a problem amenable to manual therapy (3). Confusion also exists over the effects of physical stress on the spine in causing biomechanical vertebral dysfunction (4).

One of the more common musculoskeletal complaints is neck pain, with a point prevalence of 9-14% in adults and a lifetime prevalence of approximately 33% (5-7). Age is strongly correlated with its occurrence in an ascending fashion, with the 50-59 year old age range having the highest incidence (6). It has been suggested that symptoms related to, or referred by the cervical spine are caused by injuries to the cervical spine in more than 90% of patients suffering neck pain (8). Chronic neck pain is commonly caused by whiplash injuries, in particular those resulting from motor vehicle accidents. Statistics say that 45% to 85% of patients who have suffered whiplash injuries, complain of symptoms five years after (9,10). This high rate of persisting neck pain has a direct impact upon health care costs and permanent disability (11).

Manual therapists, including chiropractors, claim that multiple treatments are required for patients with work or accident involved injuries. However, it remains to be shown when a manipulative approach is indicated and what tests or procedures best determine the need for intervention and at which stage treatment should be concluded (12,13).
Mobility testing is important in the diagnosis of back and neck pain, particularly in light of studies showing that fixation in a joint for two months or longer can lead to irreversible degenerative change (14). Motion palpation tests the “give” of musculoskeletal structures under dynamic demands (15). Dishman has suggested that without regular assessment of the biomechanical integrity of spinal motion there is an increased risk of spinal dysfunction development due to movement aberrations (16). The greatest agreement between examiners has been observed in musculoskeletal based diagnosis (17). However, there is no accepted method of measuring vertebral motion changes that has shown to be a statistically reliable. This is also true with methods of measuring vertebral motion changes utilised by chiropractors, where a motion test is often used as a diagnostic pre-requisite to mechanical intervention.

**THE CHIROPRACTIC SUBLUXATION**

Historically, manipulation has been applied to manage spinal and extremity joint lesions that produce pain (18). Early authors in the area of manipulation (19-21) believed that the cause of pain in the areas of abnormal movement of the spine was due to subluxation - a minor partial or incomplete dislocation of the vertebra. This has resulted in the term manipulable lesion, which describes a clinically significant disturbance of joint movement or position that responds to manipulation (22). This is synonymous with the entity chiropractors call the Vertebral Subluxation Complex (VSC).

The VSC is surrounded by varied opinion on its definition, however, there is one characteristic upon which there is consensus: Vertebral subluxation is dynamic and involves a restriction in normal movement of one joint interface relative to its neighbour, in one or more vectors of which it ought normally to be able (22). For the purposes of this study, the term subluxation (VSC) is defined as articular aberration producing alterations in vertebral motion. The term “chiropractic lesion” or “vertebral dysfunction” will be considered synonymous with subluxation (VSC).

**RELIABILITY OF SPINAL MOTION EXAMINATION**

Mensor and Duval (23) were probably two of the earliest researchers to examine the reliability of motion examination of the spine. They studied 527 patients with low back pain and a control group of 94 healthy individuals and found significant hypomobility of L4-5, demonstrated by stress radiographs, present in 43% of low back pain patients and in only 15% of normal individuals.

Gonella, Paris, and Kutner (24) examined segmental flexion, lateral bending and rotation of the lumbar spine in five asymptomatic female subjects in a side-lying position. Five similarly trained physical therapists with 3-20 years experience participated. Results showed a reasonably good intraexaminer reliability (although this was not quantified), and a poor interexaminer reliability. They found that greater examiner experience resulted in an even worse interexaminer reliability (again no substantiation of this statement).

In contrast with these low findings, Bergstrom and Cortis (25) found high correlation for interexaminer reliability (81.8%) for lumbar motion palpation in the seated position. Jull and Bullock (26) also show a good to high degree of interexaminer reliability for flexion, extension, rotation and lateral flexion when evaluated side-lying, based on Pearson’s correlation coefficient, (r=0.82-0.94).

The purpose of the present study was to attempt to determine whether a large group of clinicians could replicate each others findings by agreeing on the presence or absence of vertebral dysfunction in the same sample population while using their own standard clinical techniques and procedures, i.e. simulating normal everyday practice.

**METHOD**

The study involved a group of a 53 university students, all of whom were volunteers and most of whom were asymptomatic for neck pain. Most of the participants were chiropractic students at the Macquarie University's Centre for Chiropractic. Each individual underwent chiropractic examination of the cervical spine for vertebral dysfunction. Chiropractic examination involved whatever methods the examiners used in their normal everyday practice. Individual clinical methods included static and/or motion palpation, vertebral springing, range of motion and applied kinesiology. No individual had radiographs taken as part of the examination, nor were they available for comparison, even though some of the examiners normally utilise radiographs as part of their normal everyday practice.

All together there were eight examiners involved. Each examiner had an independent recorder who remained with them. The examiners clinical experience ranged from two to fourteen years (average was 8.25 years). All were graduates from either Sydney College of Chiropractic (which amalgamated with Macquarie University in 1990) or Macquarie University.
At the commencement of the study all participants were informed about the testing procedures and asked to complete a questionnaire designed for the study. This screened for exclusion criteria, which included recent neck surgery, possible pathologies and congenital deformity in the neck region. It also included a brief history on each subject indicating current or previous neck pain and associated headaches, the location of such neck pain, severity of it (by visual analogue scale), whether injury was involved and if so, what kind of injury.

The subjects then underwent analysis, were each involved and if so, what kind of injury. it (by visual analogue scale), whether injury was headaches, the location of such neck pain, severity of included a brief history on each subject indicating congenital deformity in the neck region. It also recent neck surgery, possible pathologies and This screened for exclusion criteria, which included STATISTICAL ANALYSIS

The data set collected for the examiners was analysed for two responses - a response of restricted motion or freedom of motion. Logistic regression and sequential analysis of variance using the chi-squared statistic were employed. The results were modelled, using a generalised linear model with a Bernoulli error structure and Logit link function, to find out whether there were any interaction effects between these factors. Any of these factors which proved significant in influencing the diagnosis was then further analysed for individual effect.

RESULTS

The data set for the examiners consisted of 2016 binary observations i.e. examiner analyses, (not all subjects seeing all seven of the eight examiners). There were under consideration, three possible factors influencing the analysis rendered upon any subject. These were the particular examiner carrying out the examination (exam), the subject being examined (pat) and the vertebral level being investigated (lev) which were modelled several ways. Logistic regression, including all second order interaction effects, yielded a deviance of 1335.9 on 1385 degrees of freedom (df). When patient effects were removed, so that the model was determined only by the examiner and level effects, a deviance of 1956.0 on 1960 df was obtained. Testing the difference between the two models revealed a deviance of 620.1 on 575 df with a p-value of 0.0941. This is not significant, indicating that the particular patient under examination had no bearing on the diagnosis being rendered by any examiner. Moreover, there is no relationship between the examiner and patient or the examiner and level.

ACO

Volume 5 • Number 1 • March 1996

CERVICAL SPINE PROBLEMS

A PILOT STUDY

| Source          | D.F | Deviance | p-value |
|-----------------|-----|----------|---------|
| pat*lev         | 349 | 570.2    | 0.0000  |
| exam            | 7   | 71.3     | 0.0000  |
| exam.lev        | 42  | 144.7    | 0.0000  |
| exam.pat        | 252 | 200.6    | 0.9329  |
| Residual        | 1385| 1335.9   |         |
| Total           | 2015| 2322.7   |         |

Table 1: Sequential analysis of deviance

Table 1 summarises the results of a sequential analysis of deviance or variance and provides a line by line comparison. Note that, earlier comments regarding the significance of the patient appear to be contradicted by the first row of the table. This is not so, however. The pat*lev interaction has been included first in the analysis before anything else and, as a result, appears to be significant. The table shows that the examiners are not prone to analysing particular patients in a biased manner, since exam.pat is not significant. More interesting is row three that suggests a more complex relationship between examiner and vertebral level exists and that this relationship also contributes to the result of any examination conducted. Note that the relationship between the examiner and vertebral level, while significant, is not as important as the part played by the examiner himself/herself. This conclusion is based on the observation that the deviance for exam is approximately 10 times its degrees of freedom.

| Comparison between examiners | Vertebral Level | C1   | C2   | C3   | C4   | C5   | C6   | C7   |
|------------------------------|----------------|------|------|------|------|------|------|------|
| 1 and 2                      | -3.7984        | 2.1382| 2.3303| 0.8893| 2.6148| 4.4297| 0.9794|
| 1 and 3                      | -1.4612        | 0.5246| -1.2682| 1.2676| 2.0560| 1.5276| -1.1712|
| 1 and 4                      | -3.6647        | 1.6273| 0.1543| 0.0033| 0.6992| 1.3576| -1.0675|
| 1 and 5                      | 0.9237         | 0.9560| -0.2624| 1.5671| 0.9827| 0.4371| 0.8457|
| 1 and 6                      | -2.0000        | 0.6733| 0.4652| 1.9517| 3.3535| 0.4268| -0.4001|
| 1 and 7                      | -2.8993        | 0.8851| 0.6725| 2.5619| 0.0081| 2.8220| -0.4784|
| 1 and 8                      | -1.4802        | -0.1985| -0.3043| -0.4745| -1.6890| -0.9194| -0.5977|
| 2 and 3                      | 2.8682         | -1.8021| -3.2387| -0.9293| -1.9988| -3.9777| -1.3730|
| 2 and 4                      | 1.1339         | -0.9079| -2.0785| -0.7580| -1.9318| -3.3103| -1.7625|
| 2 and 5                      | 3.0449         | -1.3295| -2.4183| 0.1671| -1.7279| -3.9278| -0.3493|
| 2 and 6                      | 1.9012         | -1.3571| -1.6398| 0.5559| 0.3242| -3.4143| -0.4522|
| 2 and 7                      | 1.3642         | -1.2981| -1.5584| 0.9776| -2.1225| 1.9249| -0.5407|
| 2 and 8                      | 3.5860         | -2.2330| -2.4733| -0.5445| -3.7031| -4.9170| -0.6612|
| 3 and 4                      | -2.4732        | 1.2078| 1.3508| -0.9036| -1.1692| -0.0623| -0.1287|
| 3 and 5                      | 0.4428         | 0.5017| 0.9722| 0.0414| -0.9287| -0.0118| 2.0312|
| 3 and 6                      | 0.8828         | 0.2716| 1.5156| 0.9543| 1.8319| -0.7406| -0.3526|
| 3 and 7                      | -1.7288        | -0.4511| 1.7938| 1.5074| -1.4922| 1.5656| -0.4216|
| 4 and 5                      | 1.2413         | -0.7102| 0.9205| -0.5383| -3.3740| -2.3727| -0.5379|
| 4 and 6                      | 2.6921         | -0.6127| -0.3978| 1.1539| 0.2464| -0.8929| 1.7589|
| 4 and 7                      | 1.1202         | 0.6809| 0.3180| 1.5225| 2.6901| -0.6712| -0.3462|
| 4 and 8                      | 0.4246         | -0.5939| 0.5006| 1.9207| -0.5392| 1.5342| -0.4140|
| 5 and 6                      | 3.3691         | -1.7644| -0.4300| -0.4741| -2.2136| -2.4816| -0.5319|
| 5 and 7                      | -1.1886        | -0.4163| 0.6702| 0.5590| 2.4290| 0.0671| -0.4335|
| 5 and 8                      | -2.0113        | 0.0159| 0.8894| 1.0599| -0.7484| 2.3440| -0.5185|
| 6 and 7                      | 0.7554         | -1.1245| -0.0406| -0.5590| -2.4605| -1.3144| -0.6386|
| 6 and 8                      | -0.6445        | 0.1243| 0.1393| 0.4005| -2.6677| 1.8529| 0.0004|
| 7 and 8                      | 1.8133         | -0.8203| -0.6963| -0.5923| -4.3234| -1.1545| -0.0690|
| 7 and 9                      | 2.6331         | -1.0265| -0.9007| -0.6175| -1.3726| -3.4725| -0.0768|

Table 2: Normal Score Comparisons between Examiners at each Vertebral Level.
Table 2 summarises, for each vertebral level, differences in the contributions made by examiners towards the result. It was derived by normalising the differences between the parameter estimates of the exam.lev second order interaction term in the final fitted model. Each number in the table may be considered normal with mean 0 and standard deviation 1. Following this, a standard hypothesis test was conducted to decide whether the difference between two examiners at a particular level was statistically significant. This data is described by Table 3. An 'X' has been placed in each position in the table where the test was significant at the p ≤0.05 level.

| Comparison between examiners | Vertebral Level | C1   | C2   | C3   | C4   | C5   | C6   | C7   |
|-----------------------------|----------------|------|------|------|------|------|------|------|
| 1 and 2                     |                | X    | X    |      |      |      |      |      |
| 1 and 3                     |                |      |      |      |      |      |      |      |
| 1 and 4                     |                |      |      |      |      |      |      |      |
| 1 and 5                     |                |      |      |      |      |      |      |      |
| 1 and 6                     |                |      |      |      |      |      |      |      |
| 1 and 7                     |                |      |      |      |      |      |      |      |
| 1 and 8                     |                |      |      |      |      |      |      |      |
| 2 and 3                     |                |      |      |      |      |      |      |      |
| 2 and 4                     |                |      |      |      |      |      |      |      |
| 2 and 5                     |                |      |      |      |      |      |      |      |
| 2 and 6                     |                |      |      |      |      |      |      |      |
| 2 and 7                     |                |      |      |      |      |      |      |      |
| 2 and 8                     |                |      |      |      |      |      |      |      |
| 3 and 4                     |                |      |      |      |      |      |      |      |
| 3 and 5                     |                |      |      |      |      |      |      |      |
| 3 and 6                     |                |      |      |      |      |      |      |      |
| 3 and 7                     |                |      |      |      |      |      |      |      |
| 3 and 8                     |                |      |      |      |      |      |      |      |
| 4 and 5                     |                |      |      |      |      |      |      |      |
| 4 and 6                     |                |      |      |      |      |      |      |      |
| 4 and 7                     |                |      |      |      |      |      |      |      |
| 4 and 8                     |                |      |      |      |      |      |      |      |
| 5 and 6                     |                |      |      |      |      |      |      |      |
| 5 and 7                     |                |      |      |      |      |      |      |      |
| 5 and 8                     |                |      |      |      |      |      |      |      |
| 6 and 7                     |                |      |      |      |      |      |      |      |
| 6 and 8                     |                |      |      |      |      |      |      |      |
| 7 and 8                     |                |      |      |      |      |      |      |      |

Table 3: Significance of Comparisons between Examiners at Each Vertebral Level. ‘X’ denotes a significant difference.

The tables reveal that examiners 3, 4, 5 and 7 are likely to arrive at similar diagnostic conclusions. In contrast, the first and second examiners differed from these four and each other. Examiners 6 and 8 could almost be placed in the homogeneous group of four, except that they differed with each other at vertebral level C5 only. Finally, the tables show that vertebral level C6 was the main area of discrepancy between examiners. C1 and C5 were next in line as points of contention between different examiners. There is essentially little or no difference between examiners at C2, C4, C5 and C7.

For vertebral level C6, examiner 2 is more likely to diagnose a problem than any of the other examiners. Examiners 1, 3, 4, 5, 6, 7 and 8 form a group in which the difference between examiners is not statistically significant. Seven of the examiners, numbers 2-7 are essentially the same in terms of their contribution towards the examination of C1. Examiner 1 stands apart from this large group, picking up significantly more problems at the C1 level.

| Vertebal Level | Examiner | Likelihood of Diagnosing a Problem Compared to the Remaining 7 examiners |
|---------------|----------|--------------------------------------------------------------------------------|
| C1            | 1        | More Likely                                                                     |
| C5            | 8        | Less Likely                                                                     |
| C6            | 2        | More Likely                                                                     |

Table 4: Vertebral Levels of Discrepancy

At the C5 vertebral level, examiner 8 stands alone from the remaining 7 as the examiner least likely to find vertebral dysfunction. At each of these levels there was a single examiner responsible for the discrepancy. Table 4 illustrates which examiners these were and how they differed from the other seven. Examiners 3-7 and examiners 3, 4, 5, 7 and 8 form relatively homogeneous groups in which there are no significant differences between any pair of examiners belonging to each group with approximate agreements of 28.57% and 30.61%, respectively.

| Examiner | Vertebral Level | Total |
|----------|----------------|-------|
|          | C1  | C2  | C3  | C4  | C5  | C6  | C7  |     |
| 1        | 26  | 21  | 11  | 5   | 11  | 12  |     | 97  |
| 2        | 5   | 9   | 5   | 1   | 6   | 15  | 3   | 44  |
| 3        | 25  | 21  | 4   | 7   | 19  | 18  | 6   | 100 |
| 4        | 15  | 17  | 4   | 1   | 5   | 8   | 2   | 52  |
| 5        | 21  | 19  | 7   | 7   | 11  | 9   | 15  | 89  |
| 6        | 12  | 7   | 4   | 5   | 14  | 4   | 0   | 46  |
| 7        | 10  | 10  | 5   | 6   | 2   | 14  | 0   | 47  |
| 8        | 21  | 17  | 8   | 0   | 3   | 6   | 0   | 55  |
|          | 135 | 121 | 48  | 32  | 71  | 85  | 38  | 530 |

Table 5: The Number of Positive Diagnoses for Examiner by Level

| Examiner | Vertebral Level | Total |
|----------|----------------|-------|
|          | C1  | C2  | C3  | C4  | C5  | C6  | C7  |     |
| 1        | 40  | 40  | 40  | 40  | 40  | 40  | 40  | 280 |
| 2        | 27  | 26  | 26  | 26  | 26  | 26  | 26  | 183 |
| 3        | 48  | 48  | 48  | 48  | 48  | 48  | 48  | 336 |
| 4        | 49  | 49  | 49  | 49  | 49  | 49  | 49  | 343 |
| 5        | 35  | 35  | 35  | 35  | 35  | 35  | 35  | 245 |
| 6        | 23  | 23  | 23  | 23  | 23  | 23  | 23  | 161 |
| 7        | 33  | 33  | 33  | 32  | 33  | 33  | 33  | 230 |
| 8        | 34  | 34  | 34  | 34  | 34  | 34  | 34  | 238 |
|          | 289 | 288 | 288 | 287 | 288 | 288 | 288 | 2016|

Table 6: The Total number of Patients Examined by each Examiner at Each Vertebral Level.
Tables 5 and 6 illustrate the number of positive findings concluded by each examiner by vertebral level and the total number of patients examined by each examiner at each vertebral level. Each number in Table 5 tells us how often each examiner found vertebral dysfunction at each vertebral level. Each number in Table 6 shows the number of times an examiner examined each vertebral level (i.e. the number of participants that the examiner observed at each level). Lastly, each entry in Table 7 represents the percentage of positive observations each examiner made at each level.

| Examiner | Vertebral Level | Total |
|----------|----------------|-------|
|          | C1  | C2  | C3  | C4  | C5  | C6  | C7  |
| 1        | 0.65| 0.53| 0.28| 0.13| 0.28| 0.28| 0.30| 0.35|
| 2        | 0.19| 0.35| 0.19| 0.04| 0.23| 0.58| 0.12| 0.24|
| 3        | 0.52| 0.44| 0.08| 0.15| 0.40| 0.38| 0.13| 0.30|
| 4        | 0.31| 0.35| 0.08| 0.02| 0.10| 0.16| 0.04| 0.15|
| 5        | 0.60| 0.54| 0.20| 0.31| 0.26| 0.43| 0.36| 0.61|
| 6        | 0.52| 0.30| 0.17| 0.22| 0.61| 0.17| 0.00| 0.29|
| 7        | 0.30| 0.30| 0.15| 0.19| 0.06| 0.42| 0.00| 0.20|
| 8        | 0.62| 0.50| 0.24| 0.00| 0.09| 0.18| 0.00| 0.23|
| Total    | 0.47| 0.42| 0.17| 0.11| 0.25| 0.30| 0.13| 0.26|

Table 7: The Percentage of Positive Diagnoses Made by Each Examiner at Each Level.

DISCUSSION

Due to the almost mythical nature of the entity known as the vertebral subluxation (VSC) and the difficulties recognised in defining it, it is hoped that the simple data provided may assist in quantifying the level of reliability which may be expected in the diagnosis of VSC. If the location of spinal dysfunction could be identified in a dependable manner by a group of skilled clinicians then perhaps one method of justifying the existence of the vertebral subluxation is inference due to consensus.

The results reveal that for all examiners there is no examiner-patient or patient-level interactions and that a significant interaction between the examiner and the vertebral level influences the outcome of the examinations. Essentially this means that, collectively, for all examiners, there was poor interexaminer reliability. However, there does appear to be combinations of examiners for which there is significant agreement. These are examiners 3-7 and examiners 3, 4, 5, 7 and 8 who agreed approximately 28.57% and 30.61% of the time, respectively.

These values may sound low and should be put into perspective in relation to other standard tests. With respiratory problems, for example, the degree of interexaminer agreement beyond chance when examining for the presence or absence of signs of airway obstruction has been measured at between 14-64% in medical physicians (27). Mior and colleagues (28) examined the cervical spines of fifty-nine subjects. Motion palpation was performed by two student examiners in their final year of clinical training (3 months intensive training) and agreement beyond chance indicated interexaminer agreement of 15%.

The largest areas of contention in the examinations were C6, C1 and C5, in descending order of controversy. It should also be noted that these areas are where most positive findings were recorded. This may be suggestive of the expectations of the examiners. They may believe that these areas are more important than others and by virtue of this belief tend to be more sensitive to minor aberrations in joint behaviour or more likely to err in favour of a positive finding if in doubt, (i.e. they may be searching for what they believe should be there as opposed to what is actually present).

PROBLEMS INHERENT IN STUDY DESIGN AND IMPLEMENTATION

As one aim of this study was to develop a protocol to investigate examiner reliability, a summary of the difficulties encountered in the study is presented.

1. The use of an asymptomatic student population. In this study, it appears that the sample of subjects possess similar vertebral characteristics (i.e. homogeneous sample). This would mean that the data set essentially represents examiner error. Johnston (29) speculates that subjects in this type of population are likely to have minor lesions that may be influenced by repetitious examination. However it has also been reported that severity of the spinal lesion does not appear to make a difference in interexaminer reliability (30).

2. Examiner selection. It was thought that using experienced practitioners would improve the replicability of what is actually being detected and that this would improve the interexaminer reliability, however, some doubt has been cast on the use of experienced clinicians. In fact, some researchers have noticed reductions in interexaminer reliability with increased practitioner experience (24,33). Others, however, indicate that experience has no effect (30).

Also, by making use of examiners trained solely at one institution, the findings cannot be related to general chiropractic practice. A future study should draw together a cohort of examiners more representative of the profession at large.
3. **Examiners using their own individual techniques for detecting areas of subluxation.** Different methods of evaluating the musculoskeletal system may also affect interexaminer reliability (31). Observations of independent examiners utilising their own systems of examination gave findings with imperfect interexaminer reliability (32). However, if the tests to be used are agreed upon prior to examination of the patient, improved examiner correlation would be expected. This has been observed by McConnell (33).

4. **Reliability of examiners in what they are detecting.** Several researchers (22,34) emphasise the difficulty in assessing a phenomenon that cannot be directly perceived since palpatory joint motion analysis takes place over several layers of tissues of varying densities. Charlton (22) also points out that if a system or method of examination is eventually shown to be reliable that this in itself does not demonstrate its validity. This reliability may be more due to the training of the examiners.

5. **Repeated examination.** Greenman (35) stresses that subjecting the musculoskeletal system to repeated diagnostic procedures may induce changes in the system. Such changes can be a source of diagnostic confusion. DeBoer et al. (36) also suggest that stress put on subjects by consecutive examination might cause subluxation.

6. **Subject Compliance.** It is noted that not all subjects were analysed the same number of times (ie. seven). This is shown by the totals listed in Table 6. This was due to the amount of time that subjects were required to participate and a lack of understanding about how often they were to be examined, although there being no comments as such made to the examiners or researchers involved.

7. **Not all examiners inspected all the patients.** Due to the time constraint on collecting the data in the one session, it was decided that subjects need see only seven of the eight examiners. However this has resulted in the statistics being somewhat incomplete, making it harder to draw conclusions.

8. ** Examiner expectation.** Do examiners expect to find certain problems or certain numbers of problems when they examine subjects? The role of prior expectation on examination results is not clear. Does it in fact have any effect and if so can training reduce it? It may be that the examiners in this study were predisposed to diagnosing vertebral levels according to a preconceived notion.

Most studies have average to low levels of interexaminer agreement, including Rhudy, Sandefur and Burk (37) who concluded that with the present lack of scientific evidence for the value of musculoskeletal diagnostic measures such as motion palpation, the judgements made by the clinician may be based more on other subjective impressions than on the information derived from the procedures themselves. The results of this study may support this idea.

Based on the statistical analyses, sufficient evidence exists to suggest a possible relationship between examiners and a prospective study should be designed to explore this possibility. It is hoped that these results will provide a basis for further work in an endeavour to reliably and validly demonstrate the vertebral subluxation complex or chiropractic lesion.

**ACKNOWLEDGMENTS**

The authors would like to thank Andrew Hart BSc (Hons), University of Queensland, for his statistical expertise.

Also, the following chiropractors who donated the time in participating in this study: Dr’s S Austin, G Baxter, J Cice, S Howard, D Montgomery, A Notenko and H Pollard.

**REFERENCES**

1. Pope MH, Andersson GBJ, Fryomoyer JW and Chaffin DB. Occupational Low Back Pain. Chicago: Mosby Yearbook, 1991.
2. Meade TW, Dwyer S, Browne W, Townsend J and Frank OA. Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment. British Medical J. 1990; 300: 1431-1436.
3. Cats-Baril WL, Fryomoyer JW. Identifying patients at risk of becoming disabled because of low back pain. Spine 1991; 16(6): 605-607.
4. Kumar S. Cumulative load as a risk factor for back pain. Spine 1989; 15(12): 1311-1316.
5. Bland JH. Disorders of the Cervical Spine. Philadelphia: Saunders, 1987: 1-8.
6. Donk J, Schouten JSA, Passchier J, Roumunde LKJ, Vankenburg HA. Traits in a general population. J Rheumatol. 1991; 18: 1884-1889.
7. Shekelle PG, Brook RH. A community-based study of the use of chiropractic services. Am. J. Public Health 1991; 81: 439-442.
8. Jackson R. The Cervical Spine 4th ed. Springfield: Charles C. Thomas, 1977.
9. Watinson A, Gargan MF, Bannister GC. Prognostic factors in soft tissue injuries of the cervical spine. Injury 1991; 22: 307-309.
10. Macnab I. Acceleration injuries of the cervical spine. J Bone Joint Surgery 1964; 46-A(8): 1797-1799.
11. Janda J. Some aspects of extracranial causes of facial pain. J Prosthet Dent. 1986; 56: 484-487.
12. Spratt KF, Lehmann TR, Weinstein JN, Sayre HA. A new approach to the low back physical examination: behavioural assessment of mechanical signs. Spine 1990; 15(2): 96-102.
13. Deyo RA. Measuring the functional status of patients with low back pain. Chiropractic Technique 1990; 2(3): 127-137.
14. Chaffin DB, Baker WJ. A biomechanical model for analysis of symmetric sagittal plane lifting. AIIE trans. 1970; 2: 16-27.
15. Herzog W, Read LJ, Conway JW, Shaw LD, McEwen MC. Reliability of motion palpation procedures to detect sacroiliac joint fixations. J Manipulative Physiol Ther. 1989; 12(2): 86-92.
16. Dishman RW. Static and dynamic components of the chiropractic subluxation complex: a literature review. J Manipulative Physiol Ther. 1988; 11(2): 98-107.
17. Leboeuf C, Gardner V. Chronic low back pain: orthopaedic and chiropractic test results. J Aust Chiro Assoc. 1989; 19: 9-16.
18. Brauntingham JW. A critical look at the subluxation hypothesis. J Manipulative Physiol Ther. 1988; 11(2): 130-132.
19. Palmer DD. The chiropractors adjustor. Portland: Portland Publishing, 1910.
20. Still AT. Osteopathy research and practice. Kirkville: Privately published, 1910.
21. Cyriax J. Lumbago: the mechanism of dural pain. Lancet 1945; 427.
22. Charlton KH. Approaches to the demonstration of vertebral subluxation: 1. Introduction and manual diagnosis: A review. J Aust Chiro Assoc. 1988; 18: 9-13.
23. Mensor MC, Duval G. Absence of motion at the fourth and fifth lumbar interspace in patients with and without low back pain. J Bone Joint Surg. 1959; 41A: 1047-54.
24. Gonnella C, Paris S, Kutner M. Reliability in evaluating passive intervertebral motion. Physical Therapy 1982; 62(4): 436-444.
25. Bergstrom E, Cortis G. An interexaminer and intraexaminer reliability study of motion palpation of the lumbar spine in lateral flexion in the seated position. European J Chiro. 1986; 34: 121-141.
26. Juli G, Bullock M. A motion profile of the lumbar spine in an ageing population assessed by manual examination. Physiotherapy 1987; 3: 70-81.
27. Godfrey S, Edwards RJ, Campbell, EJ. Repeatability of physical signs in airway obstruction. Thorax 1969; 14: 4.
28. Mior SA, King RS, McGregor M, Bernard M. Intra and interexaminer reliability of motion palpation in the cervical spine. J Canadian Chiropractic Assoc. 1985; 29(4): 195-198.
29. Johnston W. Interexaminer reliability in palpation. JAOA 1976; 76: 286.
30. Herzog W, Read LJ, Conway JW, Shaw LD, McEwen MC. Reliability of motion palpation procedures to detect sacroiliac joint fixations. J Manipulative Physiol Ther. 1989; 12(2): 86-92.
31. Beal MC, Dvorak J. Palpatory examination of the spine: A comparison of the results of two methods and their relationship to visceral disease. Manual Medicine 1984; 1: 25-32.
32. McConnell DG, Beal MC, Dinnar U, Goodridge JP, Johnston WL, Karni Z, Upledger JE, Blum G. Low agreement of findings in neuromusculoskeletal examination by a group of osteopathic physicians using their own procedures. JAOA 1980; 79: 441-450.
33. Beal MC, Goodrich JB, Johnston WL, McConnell DG. Interexaminer agreement on patient improvements after negotiated selection of tests. JAOA 1980; 79: 432-440.
34. Srachan WF. Joint motion testing and forces involved in passive motion. JAOA 1963; 66: 271-277.
35. Greenman PE. Manipulative therapy in total health care. In Korr IM, ed. The neurobiologic mechanisms in manipulative therapy. New York: Plenum Press, 1978.
36. DeBoer KF, Harmon R, Tuttle CD, Wallace H. Reliability study of detection of somatic dysfunctions in the cervical spine. J Manipulative Physiol Ther. 1985; 8(1): 9-16.
37. Rhudy, TR, Sandefur MR, Burk JM. Interexaminer/intertechnique reliability in spinal subluxation assessment: a multifactorial approach. Am J Chiro Med. 1988; 1: 111-114.