Published in final edited form as:

J Manipulative Physiol Ther. 2010 ; 33(3): 220–225. doi:10.1016/j.jmpt.2010.01.009.

Reliability of Zygapophysial Joint Space Measurements Made from MRI Scans of Acute Low Back Pain Subjects: Comparison of Two Statistical Methods

Gregory D Cramer, DC, PhD [Professor and Dean of Research]¹, Joe A Cantu, DC [Consulting Radiologist]², Judith D Pocius, MS [NUHS Research Coordinator]¹, Jerrilyn A Cambron, DC, MPH, PhD [Professor, NUHS Department of Research]¹, and Ray A McKinnis, PhD [Consulting Biostatistician, Warrenville Illinois]³

¹Department of Research, National University of Health Sciences, Lombard, Illinois

²Private Practice of Chiropractic and Radiology, Charlottesville, Virginia

³Consulting Biostatistician, Warrenville Illinois

Abstract

Objective—This purpose of this study was to assess the reliability of measurements made of the zygapophysial (Z) joint space from the MRI scans of subjects with acute low back pain (ALBP) using new equipment and 2 different methods of statistical analysis. If found reliable, the methods of Z joint measurement can be applied to scans taken before and after spinal manipulation in a larger study of ALBP subjects.

Methods—Three observers measured the central anterior-to-posterior distance of the left and right L4/L5 and L5/S1 Z joint space from 5 subject scans (20 digitizer measurements, rounded to 0.1mm) on two separate occasions separated by 4 weeks. Observers were blinded to each other and their previous work. Intra- and interobserver reliability was calculated by means of intra-class correlation coefficients (ICCs) and also by mean differences using the methods of Bland and Altman (1986). A mean difference of $\leq \pm 0.4$ mm was considered clinically acceptable.

Results—ICCs showed intraobserver reliabilities of 0.95 (95%CI 0.87-0.98), 0.83 (0.62-0.92), and 0.92 (0.83-0.96) for each of the 3 observers; and interobserver reliabilities of 0.90 (0.82-0.95), 0.79 (0.61-0.90), and 0.84 (0.75-0.90) for the first and second measurements and overall reliability, respectively. The mean difference between the first and second measurements was -0.04 mm (± 1.96 SD= $-0.37 - 0.29$), 0.23 ($-0.48 - 0.94$), 0.25 ($-0.24 - 0.75$), and 0.15 ($-0.44 - 0.74$) for each of the 3 observers and the overall agreement, respectively.

Conclusions—Both statistical methods were found to be useful and complementary and showed the measurements to be highly reliable.

© 2010 National University of Health Sciences. Published by Mosby, Inc. All rights reserved.

Please address correspondence to: Gregory D. Cramer, DC, PhD, Professor and Dean of Research, National University of Health Sciences, 200 East Roosevelt Road, Lombard, Illinois 60510, Office: (630) 889-6536, Fax: (630) 495-6664.

CONFLICTS OF INTEREST No conflicts of interest were reported for this study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Key Indexing Terms

Reliability; Manipulation, Spinal; Zygapophysial Joint; chiropractic

INTRODUCTION

Separation (gapping) of the zygapophysial (Z) joints during a spinal adjustment is thought to have a positive therapeutic affect by breaking up intra-articular adhesions.¹⁻³ In previous studies,^{4,5} we found that side-posture positioning and chiropractic manipulation gapped the lumbar Z joints in healthy volunteers as measured from magnetic resonance imaging (MRI) scans. Chiropractic manipulation resulted in greater gapping of the Z joints than did side-posture positioning. The purpose of continuing this line of investigation is to build on those findings by determining if chiropractic manipulation gaps the L4/L5 and L5/S1 Z joints (L4-S1) in subjects with acute low back pain (ALBP), and to determine if the amount of Z joint gapping is related to a decrease in subjects' pain and functional impairment. These aims can only be accomplished if the MRI measurements of ALBP patients can be made reliably.

Our research group has previously evaluated the reliability of MRI measurements taken from the lumbar Z joints, intervertebral foramina, and the vertebral canal.⁶⁻¹⁰ Studies included measurements of both cadaveric spines and living subjects using both high and low field strength MRI units, including the same MRI unit used in this study.^{7,9}

Many researchers consider properly conducted reliability studies to be an essential component of mechanisms of action and clinical research and lament the paucity of well conducted reliability studies in current research.¹¹⁻¹⁴ The reliability study of this paper was considered necessary for several reasons. First, previous lumbar spine reliability studies of the Z joints were done on the scans of young healthy subjects (less than 25 years of age).^{4,5} The current study was done on subjects with ALBP, most of whom were over 25 years of age; this cohort of patients have increased degenerative changes that can make assessment of the joint margins on MRI more challenging. Consequently, determining the reliability of measurements in this population is important. Another reason for conducting this reliability study was that new observers would be making the measurements using new measuring equipment. In addition, the data would be analyzed using the methods of Bland and Altman¹⁵ who suggest that estimating reliability by examining the distribution of the difference in repeated measures to be superior to our previous methods of assessing reliability by calculating the intra-class correlation coefficients (ICCs).¹⁶

The purpose of this study was to assess the reliability of measurements made of the Z joint space from the MRI scans of subjects with ALBP. If found reliable, the methods will be applied to Z joint measurements taken before and after spinal manipulation and before and after side posture positioning in a larger study of 112 subjects with ALBP.

METHODS

This work was approved by the National University of Health Sciences institutional review board (IRB) for protection of human research subjects; this IRB operates under CFR 45, Subsection 46. One hundred and twelve (112) subjects completed the larger (main) study. Each subject received 2 MRI scans (Hitachi MRP 5000, 0.2 Tesla MRI unit) on 2 separate occasions (initial scans and after 2 weeks of chiropractic care) for a total of 448 films. The first scan of each MRI appointment was taken in the neutral (supine) position and the second was taken following an intervention (side posture spinal adjusting, side posture positioning, no intervention control). Z Joint measurements were to be made of the left and right L4/L5 and

L5/S1 levels (4 measurements per scan) of the 448 scans in the study if the results of the reliability study indicated the measurements could be made reliably. The study radiologist (JC) chose the specific image to be measured for each Z joint from 5 images of each segmental level using a rigorous method designed to identify the image that demonstrated the Z joint space to best advantage. This method was used in previous studies⁴ and for the 112 subjects of the larger study. Those images selected for each of the four levels were magnified at 2X and printed together on one sheet of 14 in x17 in x-ray film. The MRI scans were coded using random numbers so that all investigators (including the radiologist) had no knowledge of whether the scan was from the first or second MRI appointment or if the scan was the first or second taken during an appointment. Five subject scans (20 Z joints) were randomly selected for use in this reliability study.

Three observers were chosen from students enrolled in a complementary and alternative medicine (CAM) professional program at the National University of Health Sciences. The students had completed the first year gross anatomy course, including spinal anatomy. Using gross anatomical sections and MRI scans that corresponded to those sections, the students received additional tutoring in cross sectional anatomy of the spine from an anatomist specializing in the spine (GC).¹⁷ Particular attention was paid to instruction regarding the cross sectional anatomy of the Z joints and their appearance on MRI.

Training Protocol for Observers

The primary radiologist of the project (JC) had served in the same capacity in the previous studies. He chose two representative patient scans from the project to be used in a tutorial for the observers' measurements. The 3 observers then met with the radiologist who demonstrated the measurements on the two scans (8 measurements, 2 scans each with 4 Z joints). The measurement made of each Z joint was the shortest anterior-posterior (A-P) distance between the superior and inferior articular processes at the center of the Z joint space (Figure 1). More specifically, each measurement began from the point of the superior articular facet closest to a point bisecting a line passing between the medial and lateral extremes of the joint. Measurements began and ended at the point of markedly low signal intensity adjacent to the joint space, passing through the region of intermediate signal intensity sometimes associated with Z joint spaces as seen on MRI scans. Measurements were made on a GTCO Calcomp Drawing Board III (backlit) digitizer (Source Graphics, Anaheim, CA) and points digitized using the tablet were converted to distance by Excel Distance/Length Digi digitizing software (Logic Group, Austin, TX). The observer measured each joint space 5 times. If 3 of the 5 measurements were within 0.1mm of each other, the 3 values were averaged and recorded electronically. If 3 of 5 measurements did not agree, the measurements were repeated until 3 measurements within 0.1mm were attained. Once the observers expressed a clear understanding of the measurements, each scheduled a time to make them alone during the following week. After the 3 observers had completed their "solo," each met separately a second time with the radiologist. During this session, the radiologist discussed any difficulties the observer had experienced and the radiologist worked to clarify the observer's understanding of the measurements. The observer then made another set of measurements under the observation of the radiologist to ensure that the observer was measuring according to the previous instructions. This process continued until the radiologist was satisfied that the observer was ready to begin the reliability study.

Reliability Study

Utilizing the 5 images chosen by the radiologist (different from the scans used in the training sessions), the observers made the 20 L4/L5 and L5/S1 Z joint measurements from the subject scans on two separate occasions, separated by at least 4 weeks. The observers had no access to the measurements of one another or their previous measurements.

Data was analyzed by calculating intra-class correlation coefficients (ICCs).¹⁶ For intra-observer reliability, ICCs were calculated comparing the measurements of the first and second sets of measurements of each observer. Inter-observer reliabilities were calculated using ICCs of all three observers for the measurements of the first measurement session, second measurement session, and for the means of measurements of first and second sessions for each observer (overall reliability). ICCs assess all three observers simultaneously.

Reliability was also calculated using the methods of Bland and Altman,¹⁵ who suggest that the reliability be estimated by examining the distribution of the difference in repeated measures. Consequently, for each observer (intra-observer reliability) the measures recorded from each Z joint during the first 3 measures (first session) were subtracted from the mean of the second three (second session). The resulting values were graphically plotted against the average of the means of the first and second session measurements. In addition, the overall mean difference between the two sets of measurements was also calculated. These methods were also used for the pooled data of the three observers for each measurement session (inter-observer reliability). Acceptable reliability was set at a mean difference less than the absolute value of 0.4 mm; the value determined by the investigative team to be the minimum clinically relevant difference that could be assessed by MRI.

Results

The training and measurement protocols were all successfully completed with two separate sets of measurements, separated by a minimum of 4 weeks, completed by the 3 observers.

The values for the measurements made by the three observers are shown in Table 1. ICCs comparing the measurements of the first and second sets of measurements of each observer (intra-observer reliabilities) were: 0.95 (95%CI: 0.87-0.98), 0.83 (0.62-0.92), and 0.92 (0.83-0.96) for each of the 3 observers. Comparisons of the measurements of the three different observers (inter-observer reliabilities) were: 0.90 (0.82-0.95), 0.79 (0.61-0.90), and 0.84 (0.75-0.90) for the measurements of the first measurement session, second measurement session and overall reliability (means of measurements of first and second sessions for each observer). ICCs assess all three observers at once.

The Bland and Altman (1986) method for assessing the mean difference between the first and second measurements resulted in values of -0.04 mm (± 1.96 SD= -0.37 – 0.29), 0.23 (-0.48 – 0.94), 0.25 (-0.24 – 0.75), and 0.15 (-0.44 – 0.74) for each of the 3 observers and the overall agreement, respectively. A mean difference of $\leq \pm 0.4$ was considered clinically acceptable. Figure 2 shows the Bland and Altman plots¹⁵ of the differences between the first and second sets of measurements plotted against the mean of the same values.

DISCUSSION

Assessment of Methods

The MRI and measurement equipment performed well. All methods, including the observer measurement protocols, were successfully conducted. Two different methods of statistical analysis were used in this study to assess the intra-observer and inter-observer reliability of measurements made of the Z joint from the MRI scans of subjects with acute low back pain. The ICCs give a point estimate of the reproducibility of a measurement.¹⁶ The ICCs for this study, estimating the reliability of repeated measures by the same observer and of measures by different observers of the same joint, were very high, indicating that the stability of this evaluation process is quite adequate for research purposes.

The Bland and Altman method for estimating reliability provides a way of quantifying the variability of the process, so that it can be compared with any clinically significant difference.

¹⁵ With a difference of $<\pm 0.4$, the results of this study showed that intra- and inter-observer variability was quite good and would be adequate for clinical studies.

Observer 1 performed better than the other two observers. We believe this observer's results are a reflection of his highly meticulous approach to the measurements. Observation of this individual's measurement style by the investigators revealed that Observer 1 had a very careful approach and spent more time making individual measurements.

Limitations

One potential limitation of the study is that 20 measurements were made. A higher number of measurements might seem more desirable. However, a higher number of measurements would increase the likelihood of achieving higher reliability coefficients. Stronger results with the lower number of measurements of this study indicate that measurements could be made with a very high degree of reliability. Additional measurements are not needed to draw this conclusion. The results reflect those of the three observers who were a part of this study. Different results might be found with three different observers. The observers of this study were highly trained and comparable results would be difficult to repeat by randomly choosing untrained students or clinicians. However, with similar observer training, the results of this study should be repeatable with any three students or clinicians with a background similar to that of the observers used in this study.

CONCLUSION

Both statistical methods provided important and complimentary information. The results of both the ICCs and mean differences (i.e., Bland and Altman method) showed excellent reliability. We conclude that the measurement methods used in this study can be applied to future research assessing the Z joint space.

PRACTICAL APPLICATIONS

- Reliability studies of clinical trial outcome measures are essential
- A reliability study, using two established but quite different methods of assessing reliability, was conducted on MRI measurements made of lumbar zygapophysial (Z) joint spaces (L4/L5 and L5/S1)
- The two reliability methods provided complimentary information
- The Z joint measures were found to be reliable

Acknowledgments

We thank Frank Balester, MSOM, Lac, Derek Simpson, Tyra Horner, and Joshua Healy, DC for their help with this manuscript.

FUNDING SOURCES This study was supported by the National Institutes of Health/National Center for Complementary and Alternative Medicine (Grant # 2R01 AT000123).

References

1. Janse, J. Principles and practice of chiropractic an anthology. Wheaton, IL: Kjellberg & Sons, Inc; 1976.
2. Mooney V, Robertson J. The facet syndrome. Clin Orthop Res 1976;115:149–56.
3. Triano, J. Interaction of spinal biomechanics and physiology. In: Haldeman, S., editor. Principles and practice of chiropractic. 2. East Norwalk, Conn: Appleton & Lange; 1992.

4. Cramer G, Tuck N, Knudsen J, et al. Effects of side-posture positioning and side-posture adjusting on the lumbar zygapophyseal joints as evaluated by magnetic resonance imaging: a before and after study with randomization. *J Manipulative Physiol Ther* 2000;23:380–94. [PubMed: 10951308]
5. Cramer G, Gregerson D, Knudsen J, Hubbard B, Ustas L, Cantu J. The effects of side-posture positioning and spinal adjusting on the lumbar Z joints: a randomized controlled trial with sixty-four subjects. *Spine* 2002;27(22):2459–66. [PubMed: 12435975]
6. Cramer G, Cantu J, Dorsett R, et al. Dimensions of the lumbar intervertebral foramina as determined from the sagittal plane magnetic resonance imaging scans of 95 normal subjects. *J Manipulative Physiol Ther* 2003;26:160–70. [PubMed: 12704308]
7. Cramer G, Cantu J, Greenstein J, et al. Oblique MRI of the cervical intervertebral foramina: a comparison of three techniques. *J Neuromusculoskelet Syst* 2002;10:41–51.
8. Dorsett, R.; Cramer, G.; Howe, J., et al. Lumbar vertebral canal dimensions of eighty-eight normal subjects evaluated by 0.35 tesla magnetic resonance imaging. In: Rosner, T., editor. *Proceedings of the 1994 International Conference on Spinal Manipulation: Proceedings of the International Conference on Spinal Manipulation; 1994 June 10-11; Palm Springs, USA. Arlington: Foundation for Chiropractic Education and Research; 1994. p. 96-8.*
9. Greenstein J, Cramer G, Howe J, et al. Comparison of 1.5 tesla and 0.35 tesla field strength magnetic resonance imaging scans in the morphometric evaluation of the lumbar intervertebral foramina. *J Manipulative Physiol Ther* 1995;18:195–202. [PubMed: 7636408]
10. Cramer G. Comparison of computed tomography to magnetic resonance imaging in the evaluation of the lumbar intervertebral foramina. *Clin Anat* 1994;7:173–80.
11. Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. *J Am Med Inform Assoc* 1999 Mar;6(2):143–50. [PubMed: 10094067]
12. Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials* 2004;1(6):553–66. [PubMed: 16279296]
13. Hartmann DP. Considerations in the choice of interobserver reliability estimates. *J Appl Behav Anal* 1977;10(1):103–16. [PubMed: 16795538]
14. Haas M. How to evaluate intraexaminer reliability using an interexaminer reliability study design. *J Manipulative Physiol Ther* 1995 Jan;18(1):10–5. [PubMed: 7706953]
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 Feb 8;1(8476):307–10. [PubMed: 2868172]
16. Shrout PE, Fleiss JL. Intraclass correlations. Uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8. [PubMed: 18839484]
17. Cramer, G.; Darby, S. *Basic and clinical anatomy of the spine, spinal cord, and ANS*. Second. St Louis: Elsevier/Mosby; 2005.

Figure 1. Illustration (A) and MRI scan (B) showing the central anterior to posterior (A-P) measurement of the zygapophysial (Z) joints that were made from the left and right L4/L5 and L5/S1 Z joints in this study.

Figure 2.

Bland and Altman (1986) plots showing the differences (Y axis) between Measurement 1 and Measurement 2 (separated by a minimum of 4 weeks) plotted against the averages (X axis) of the same two measurements. The plots include the values of Observers 1-3 (A-C) and the values of all three observers to assess overall agreement (Fig. 2D).

Table 1

Measurements of Zygapophysial (Z) Joint Space from MRI Scans

Scan/Level	Measurement 1*			Measurement 2		
	Obs 1 [†]	Obs 2	Obs 3	Obs 1	Obs 2	Obs 3
1/R [‡] L4-5	4.41	3.97	4.41	4.26	3.57	3.86
1/L L4-5	3.98	3.47	3.86	3.98	4.15	3.89
1/R L5-S1	3.50	3.37	3.02	3.68	3.14	3.06
1/L L5-S1	3.64	3.91	4.26	3.69	3.28	4.22
2/R L4-5	3.58	3.60	3.68	3.81	3.32	3.49
2/L L4-5	4.37	4.59	4.34	4.34	4.20	3.89
2/R L5-S1	4.14	4.53	4.00	4.16	4.07	3.95
2/L L5-S1	4.20	3.86	4.21	4.07	3.62	4.18
3/R L4-5	4.35	3.48	4.37	4.18	3.16	3.54
3/L L4-5	4.08	3.78	3.81	4.40	2.79	3.72
3/R L5-S1	4.34	4.28	4.51	4.47	4.12	4.38
3/L L5-S1	4.43	4.33	4.70	4.23	3.79	4.42
4/R L4-5	4.20	4.19	4.42	4.24	4.15	4.12
4/L L4-5	4.29	4.46	4.32	4.03	4.81	4.32
4/R L5-S1	3.36	3.95	3.61	3.68	3.69	3.53
4/L L5-S1	3.82	3.86	3.94	3.84	4.00	3.65
5/R L4-5	3.18	3.29	3.43	3.22	2.88	3.16
5/L L4-5	3.35	3.44	3.33	3.31	3.60	2.97
5/R L5-S1	3.35	3.50	3.53	3.63	3.30	3.12
5/L L5-S1	3.49	3.49	3.78	3.58	3.15	3.00

* Measurements 1 and 2 were separated by a minimum of 4 weeks;

[†] Observers were blinded to one another and had no access to their previous measurements

[‡] L=left, R=right