J. Phys. Ther. Sci. 28: 2210–2213, 2016

The Journal of Physical Therapy Science

Original Article

Absolute and relative reliability of lumbar interspinous process ultrasound imaging measurements

Ryosuke Tozawa[1, 2)*], Munenori Katoh[1)], Hidefumi Aramaki[1)], Tsubasa Kawasaki[1)], Yuichi Nishikawa[3)], Tsuneo Kumamoto[2)], Osamu Fujinawa[2)]

[1)] *Department of Physical Therapy, Faculty of Health Science, Ryotokuji University: 5-8-1 Akemi, Urayasu, Chiba 279-8567, Japan*
[2)] *Graduate Course of Health and Social Services, Saitama Prefectural University, Japan*
[3)] *Department of Clinical Neuroscience and Therapeutics, Graduate School of Biomedical and Health Sciences, Hiroshima University, Japan*

Abstract. [Purpose] The intra- and inter-examiner reliabilities of lumbar interspinous process distances measured by ultrasound imaging were examined. [Subjects and Methods] The subjects were 10 males who had no history of orthopedic diseases or dysfunctions. Ten lumbar interspinous images from 360 images captured from 10 subjects were selected. The 10 images were measured by nine examiners. The lumbar interspinous process distance measurements were performed five times by each examiner. In addition, four of the nine examiners measured the distances again after 4 days for test-retest analysis. In statistical analysis, the intraclass correlation coefficient was used to investigate relative reliability, and Bland-Altman analysis was used to investigate absolute reliability. [Results] The intraclass correlation coefficients (1, 1) for intra-examiner reliability ranged from 0.985 to 0.998. For inter-rater reliability, the intraclass correlation coefficient (2, 1) was 0.969. The intraclass correlation coefficients (1, 2) for test-retest reliability ranged from 0.991 to 0.999. The Bland-Altman analysis results indicated no systematic error. [Conclusion] The results indicate that ultrasound measurements of interspinous process distance are highly reliable even when measured only once by a single person.

Key words: Ultrasound imaging, Lumbar interspinous process, Reliability

(This article was submitted Mar. 5, 2016, and was accepted Apr. 26, 2016)

INTRODUCTION

In patients with low back pain (LBP), hypermobility and hypomobility of the lumbar spine are often observed clinically[1)]. Evaluation of the mobility of the lumbar spine is important because it has been suggested that abnormal mobility is associated with LBP[2)]. Currently, clinical measurement techniques of global lumbar spine motion, such as inclinometry, measurement of fingertip-to-floor distance, and the modified Schober technique (MST), are used[3, 4)]. MST has been shown to have high validity and reliability[4, 5)]; therefore, clinicians regard MST as the leading evaluation technique in clinical settings. However, MST cannot express the segmental mobility of the lumbar spine. Clinically, segmental mobility is evaluated using palpation, but the reliability of palpation is not high[6)]. Other evaluation methods include computed tomography, magnetic resonance imaging (MRI), and X-ray radiography[7–9)]. These imaging methods are not frequently used because of their exposure risks and high costs in clinical settings.

In this study, evaluations of lumbar mobility using ultrasound (US) imaging, which has been used recently in musculoskeletal evaluations, were performed[10, 11)]. Chleboun et al.[12)] reported the reliability and validity of US method for the

*Corresponding author. Ryosuke Tozawa (Email: tozawa@ryotokuji-u.ac.jp)

measurement of the interspinous process distance of the lumbar spine. Their results showed that the US imaging technique was highly reliable and valid. However, they used a custom-made bed with a hole in its center, to allow comparison of the US results with MRI results, and the positions during examinations [supine on the table (lumbar spine in a neutral position), supine with lumbar, hip and knee joint flexion, and supine with lumbar extension created by inserting a foam wedge between the lumbar spine and bed in the supine position] were different from those typically used in clinical practice. Accordingly, the clinical application of the US method is problematic, because the method requires a custom-made bed, and the measurements need skill to adequately adjust the probe from under the bed. Furthermore, the top of the spinous process was used as an indicator point for measurements. Because the shape of the lumbar spinous process is wide and round, it can be difficult to define the top of the process, which may affect the reliability of measuring the distances of the interspinous processes, and measurement techniques that have high reliability and validity are needed for clinical use.

A method for quantitatively evaluating lumbar mobility at the segmental level using US diagnostic apparatus that is expected to be clinically useful was invented[13)]. Three postures were verified in that study, which were different from those of the preceding study, because these postures are easy to assess in clinical practice compared to the postures used in the preceding study[12)]. The purpose of this study was to evaluate the reliability of the novel US method of measuring the interspinous process distances of the lumbar spine.

SUBJECTS AND METHODS

The subjects were 10 males who had no history of orthopedic diseases or dysfunctions. The subjects' mean ± standard deviation (SD) age, height, weight, and body mass index were 21.7 ± 1.1 years, 170.7 ± 4.1 cm, 60.4 ± 5.7 kg, and 20.7 ± 1.1 kg/m^2, respectively. US images of the subjects' lumbar spines were acquired in the following three postures: prone lying, puppy posture, and kneel sitting with the lumbar spine fully flexed. The US imaging was performed by an experienced physical therapist with 2 years of clinical experience. A linear US probe of 6.3 MHz ± 20% (8L-RS; GE Healthcare Co.) was used with the US diagnostic apparatus (Vivid-I; GE Healthcare Corp.). The probe was placed perpendicular to the body surface and positioned parallel to the lumbar vertebrae (L1–2, L2–3, L3–4, and L4–5) to acquire images of the interspinous process. The US images were captured three times at each segment for a total of 360 images (Fig. 1). For the reliability analysis, the US images of the L1–2 interspinous process were used because they were the clearest of the interspinous processes for levels L1–2–L4–5. Also, the first results of the three trials were used because there were no major differences among the US images. Ten images were chosen at random from the 30 images to eliminate bias due to posture and subject. The 10 images used for measurements of the distance of the interspinous process were performed by nine physical therapy students. Four test–retest examiners were selected from among the nine examiners to perform the measurements again after 4 days. The measurement of the interspinous process was performed between the caudal end of L1 and the cranial end of L2 (Fig. 2). The examiners were give verbal instructions, such as: "The top of the spinous looks like a white mountain. Please measure the distance between the beginning and end of the mountain." In addition, it was ensured that the measurer was unable to see the results during measurement. The lumbar interspinous process measurements were performed five times by the examiners.

The intraclass correlation coefficient (ICC) of the nine examiners was calculated to determine the relative reliability. Inter-examiner reliability was performed using the ICC (2, 1) method, and intra-examiner reliability was performed using the ICC (1, 1) method. ICC (1, 2) was used for the results of the test-retest examinations. The analyses were performed using the statistical analysis software, R2.8.1. Examinations of absolute reliability were performed using Bland-Altman analysis[14, 15)] for the test-retest results.

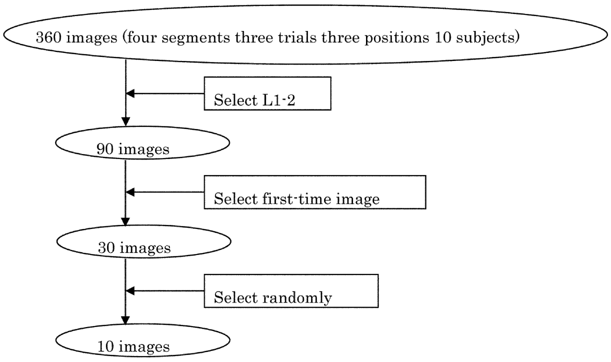

Fig. 1. Selection method of captured US images

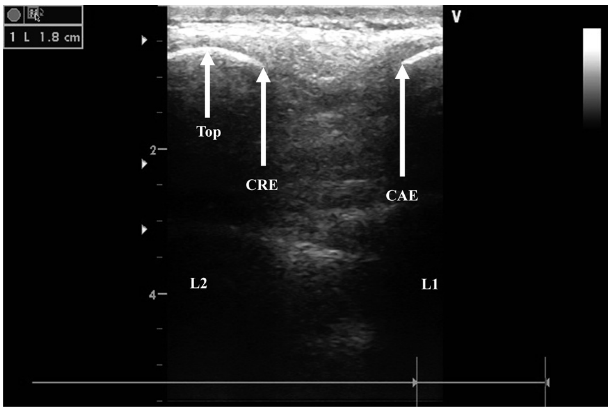

Fig. 2. Method of measuring the caliper and L1–2 interspinous process image

CRE: cranial end, CAE: caudal end, Top: spinous top

This research was approved by the ethics committee of Saitama Prefectural University (Approval number 24713), and all subjects provided their written and oral informed consent.

RESULTS

The ICCs (1, 1) of the nine examiners' measurements ranged from 0.985 to 0.998 for intra-examiner reliability, and ICC (2, 1) was 0.969 for the inter-examiner reliability (Table 1). The results of the retest were compared with those of the first measurement by analyzing the results of ICC (1, 1). ICCs (1, 2) of the measurements in the test–retest examinations were in the range of 0.991–0.999 for intra-examiner reliability (Table 2). In the results of the Bland–Altman analyses, proportional and fixed errors were not detected in the measurements. The standard error measurements ranged from 0.224–0.650 mm, and the 95% confidence intervals of minimal detectable change (MDC_{95}) ranged from 0.620–1.801 mm (Table 3).

DISCUSSION

This was a preliminary study of a US quantitative lumbar evaluation method at the segmental mobility level. Intraclass correlation coefficients between 0.41 and 0.60 characterize reliability as a "moderate", between 0.61 and 0.80 as "substantial", and ≥0.81 as an "almost perfect"[16)]. In the present study, the ICCs of relative reliability were >0.985 for ICC (1, 1) and 0.969 for ICC (2, 1). The test-retest ICC (1, 2) values were ≥0.991. All ICCs that assessed intra- and inter-examiner reliabilities were at the "almost perfect" level. In addition, systematic error was not detected in the absolute reliability results of the Bland-Altman analysis. MDC_{95} (i.e., random error in the measurement) ranged from 0.620 to 1.801 mm. For a subject, changes <MDC_{95} cannot be reliably (with a confidence level of 95%) distinguished as "real" changes in the value from chance fluctuations in the value[17)]. However, MDC_{95} was ≤1.801 mm for the interspinous process distance measurement on US images. Therefore, there is a possibility that the measurement accuracy reduced the reliability of measuring the overall process using the US diagnostic apparatus. One of the reasons that MDC_{95} increased is that there was some ambiguity in orienting landmarks before the measurements. Regarding the orientation of landmarks, only verbal commands were used in this study. It may be necessary to provide detailed landmark orientation instructions to lower MDC_{95}.

A high ICC value with no systematic error was observed for measurement of interspinous process distances on US

Table 1. Intra- and inter-examiner reliabilities of the lumbar interspinous process measurements

Intra-examiner reliability		
Measurer	ICC (1, 1)	95% Confidence interval
A	0.997	0.993–0.999
B	0.992	0.981–0.998
C	0.998	0.996–0.999
D	0.985	0.963–0.996
E	0.991	0.978–0.997
F	0.995	0.987–0.998
G	0.995	0.987–0.999
H	0.992	0.980–0.998
I	0.990	0.977–0.997
Inter-examiner reliability		
Number of measurers	ICC (2, 1)	95% Confidence interval
n=9	0.969	0.896–0.996

Table 2. Intra-examiner reliability of the test–retest

Measurer	ICC (1, 2)	95% Confidence interval
D	0.991	0.993–0.999
F	0.996	0.981–0.998
H	0.999	0.963–0.996
I	0.993	0.978–0.997

Table 3. SEM, MDC_{95}, and results of the Bland–Altman analysis between the test and retest

Measurer	Fixed bias		Proportional bias		SEM	MDC_{95}
	95% Confidence interval	Result	Regression line	Result	(mm)	(mm)
D	−0.857–0.457	no	0.012 p=0.867	no	0.650	1.801
F	−0.506–0.306	no	0.009 p=0.838	no	0.401	1.113
H	−0.126–0.326	no	−0.013 p=0.613	no	0.224	0.620
I	−0.364–0.764	no	0.041 p=0.509	no	0.558	1.546

images. This result indicates that measurements of interspinous process distances on US images can be highly reliable even if measured only once by a single person. However, a high random error of 1.801 mm was observed, which indicates that some improvements, such as definition of indicator points, are required for future measurements. Based on the results in this study, the reliability of the US method for quantitative lumbar mobility evaluations of segmental levels was demonstrated.

REFERENCES

1) Fritz JM, Whitman JM, Childs JD: Lumbar spine segmental mobility assessment: an examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehabil, 2005, 86: 1745–1752. [Medline] [CrossRef]
2) Elnaggar IM, Nordin M, Sheikhzadeh A, et al.: Effects of spinal flexion and extension exercises on low-back pain and spinal mobility in chronic mechanical low-back pain patients. Spine, 1991, 16: 967–972. [Medline] [CrossRef]
3) Miller SA, Mayer T, Cox R, et al.: Reliability problems associated with the modified Schöber technique for true lumbar flexion measurement. Spine, 1992, 17: 345–348. [Medline] [CrossRef]
4) Gill K, Krag MH, Johnson GB, et al.: Repeatability of four clinical methods for assessment of lumbar spinal motion. Spine, 1988, 13: 50–53. [Medline] [CrossRef]
5) Rezvani A, Ergin O, Karacan I, et al.: Validity and reliability of the metric measurements in the assessment of lumbar spine motion in patients with ankylosing spondylitis. Spine, 2012, 37: E1189–E1196. [Medline] [CrossRef]
6) Hestbaek L, Leboeuf-Yde C: Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther, 2000, 23: 258–275. [Medline] [CrossRef]
7) Svedmark P, Tullberg T, Noz ME, et al.: Three-dimensional movements of the lumbar spine facet joints and segmental movements: in vivo examinations of normal subjects with a new non-invasive method. Eur Spine J, 2012, 21: 599–605. [Medline] [CrossRef]
8) Xia Q, Wang S, Passias PG, et al.: In vivo range of motion of the lumbar spinous processes. Eur Spine J, 2009, 18: 1355–1362. [Medline] [CrossRef]
9) Johnsen LG, Brinckmann P, Hellum C, et al.: Segmental mobility, disc height and patient-reported outcomes after surgery for degenerative disc disease: a prospective randomised trial comparing disc replacement and multidisciplinary rehabilitation. Bone Jt J, 2013, 95-B: 81–89. [Medline] [CrossRef]
10) Brancaccio P, Somma F, Provenzano F, et al.: Changes in muscular pennation angle after crenotherapy. Muscles Ligaments Tendons J, 2013, 3: 112–115. [Medline]
11) Yu JY, Jeong JG, Lee BH: Evaluation of muscle damage using ultrasound imaging. J Phys Ther Sci, 2015, 27: 531–534. [Medline] [CrossRef]
12) Chleboun GS, Amway MJ, Hill JG, et al.: Measurement of segmental lumbar spine flexion and extension using ultrasound imaging. J Orthop Sports Phys Ther, 2012, 42: 880–885. [Medline] [CrossRef]
13) Tozawa R, Katoh M, Aramaki H, et al.: Reliability and validity of an ultrasound-based imaging method for measuring interspinous process distance in the lumbar spine using two different index points. J Phys Ther Sci, 2015, 27: 2333–2336. [Medline] [CrossRef]
14) Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986, 1: 307–310. [Medline] [CrossRef]
15) Ludbrook J: Comparing methods of measurements. Clin Exp Pharmacol Physiol, 1997, 24: 193–203. [Medline] [CrossRef]
16) Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics, 1977, 33: 159–174. [Medline] [CrossRef]
17) Faber MJ, Bosscher RJ, van Wieringen PC: Clinimetric properties of the performance-oriented mobility assessment. Phys Ther, 2006, 86: 944–954. [Medline]