Reliability and validity of an ultrasound-based imaging method for measuring interspinous process distance in the lumbar spine using two different index points

RYOSUKE TOZAWA, RPT, MS1-3)*, MUNENORI KATOH, RPT, PhD3), HIDEFUMIARAMAKI, RPT, MS1), TSUNEOKUMAMOTO, RPT, PhD3), OSAMUFUJINAWA, RPT, PhD, OMT3)

1) Department of Physical Therapy, Faculty of Health Science, Ryotokuji University: 5-8-1 Akemi, Urayasu, Chiba 279-8567, Japan
2) Department of Physical Therapy, Kasai Clinic of Orthopedic and Internal Medicine, Japan
3) Graduate Course of Health and Social Services, Saitama Prefectural University, Japan

Abstract. [Purpose] This study assessed the reliability and validity of an ultrasound-based imaging method for measuring the interspinous process distance in the lumbar spine using two different index points. [Subjects and Methods] Ten healthy males were recruited. Five physical therapy students participated in this study as examiners. The L2–L3 interspinous distance was measured from the caudal end of the L2 spinous process to the cranial end of the L3 spinous process (E-E measurement) and from the top of the L2 spinous process to the top of the L3 spinous process (T-T measurement). Intraclass correlation coefficients were calculated to estimate the relative reliability. Validity was assessed using a model resembling the living human body. [Results] The reliability study showed no difference in intra-rater reliability between the two measurements. However, the E-E measurement showed higher inter-rater reliability than the T-T measurement (Intraclass correlation coefficients: 0.914 vs. 0.725). Moreover, the E-E measurement method had good validity (Intraclass correlation coefficients: 0.999 and 95% confidence interval for minimal detectable change: 0.29 mm). [Conclusion] These results demonstrate the high reliability and validity of ultrasound-based imaging in the quantitative assessment of lumbar interspinous process distance. Of the two methods, the E-E measurement method is recommended.

Key words: Ultrasound imaging, Interspinous process distance, Reliability

INTRODUCTION

In patients with low back pain (LBP), hypermobility and hypomobility of the lumbar spine are often clinically evident1). Evaluation of lumbar spine mobility is important as abnormal mobility has been suggested to be associated with LBP3). Traditionally, quantitative evaluation methods of lumbar spine mobility have included range of motion assessment, the finger-floor distance test, and the modified Schober test. However, these methods cannot be performed at the lumbar segmental levels. In clinical practice, lumbar spine mobility evaluations at the segmental level are commonly performed by palpation; however, this method does not provide quantitative data. Manual (palpation) or visual tests that fulfill minimum consistency, reliability, and validity criteria have not been identified yet3).

Recently, ultrasound (US) imaging has been used for musculoskeletal evaluations4, 5). Chleboun et al.6) reported an US-based imaging method for measuring interspinous process distance in the lumbar spine. Their results established the reliability of US imaging. In their study, a custom bed with a hole in its center was used, and the findings from US imaging were compared with those from magnetic resonance imaging. The body position required for this examination was different from the regular positions adopted by the patients in a clinical setting; therefore, clinical application of these results was difficult. Furthermore, this previous study used the top of the spinous process as an index point for measurements. The lumbar spinous processes are wide and round; it is therefore difficult to define the top of a lumbar spinous process. This difficulty may affect the reliability of the measurement method.

Therefore, this was a pre-clinical study on an US-based imaging method used for evaluating lumbar spine mobility. The purpose of this study was to assess the reliability and validity of two US-based imaging methods in the measurement of the interspinous process distance in the lumbar spine. The reliability study compared differences in interspinous process distances between two measurement methods (different index points) using US imaging of the L2–3 interspinous
processes in humans. The validity study used index points based on the reliability study results. In addition, the validity study was carried out using a model resembling the living human body (MRLB).

**SUBJECTS AND METHODS**

**Subjects**

Ten men who had no history of orthopedic diseases or dysfunctions were selected for the reliability study. The mean (standard deviation) 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, respectively. The validity study was carried out using an MRLB. MRLB was built using a plastic skeletal model, paper towels, US gel, and agar gel. First, the plastic skeletal model was placed on a bed in the prone position. The intervertebral foramina were blocked with paper towels to prevent leakage of the US gel. Next, the interspinous process regions were filled with a sufficient amount of US gel. Finally, the tips of the spinous processes were covered with an agar gel. The agar gel was made using 400 g of agar powder and 500 mL of purified water (KENEI Pharmaceutical Co. Ltd) and mixed carefully to prevent the formation of air bubbles.

**Methods**

In the reliability study, the measurements were performed using US images of the lumbar spine in three different postures: prone, prone on elbow, and kneeling with the lumbar spine fully flexed. Measurements were taken by a physical therapist with two years of clinical experience. An US inspection apparatus (GE Healthcare Corp., Vivid-i) and a 6.3 MHz ± 20% linear US probe (GE Healthcare Corp., 8L-RS) were used for US imaging. The probe was placed perpendicular to the body surface and in a position parallel to a line connecting the lumbar spinal levels (L1–2, L2–3, L3–4, and L4–5) to collect images of the spinous process. Three US images were taken at each level. The L2–3 spinal level was selected for this study as the top of its interspinous region appeared in 17 of the 30 images. Five physical therapy students participated in this study as examiners. Prior to the measurement, the physical therapist was verbally instructed regarding the measurement procedure. All examiners received the same instructions. Two methods were used to measure the interspinous process distances in the 17 images (L2–3 interspinous process; Fig. 1). One method measured the distance between the caudal end of the L2 spinous process and the cranial end of the L3 spinous process (E-E measurement). The other method measured the distance between the top of the L2 spinous process and the top of the L3 spinous process (T-T measurement). The physical therapist recorded the measured values, and the examiner was blinded to these values. Intraclass correlation coefficients (ICCs) were calculated for each measurement to estimate the relative reliability. Inter- and intra-measurement reliability were estimated using ICC (2, 1) and ICC (1, 1), respectively.

Furthermore, the validity was examined based on the results of the reliability study. The interspinous process distance in the MRLB was measured using two methods (Fig. 2): US imaging (US method) and Spearman esthesiometer (SE method). The US method used the same device and methods as those in the reliability study. After the US method, the SE method was used to measure the interspinous process distances in the plastic skeletal model. Interspinous process distance was measured two times with each method at each segmental level. Highest precision measurements were obtained in both cases (US method, 1 mm; SE method, 0.1 mm).

ICCs were calculated for each measurement to estimate relative reliability. Inter- and intra-measurement reliability were estimated using ICC (2, 1) and ICC (1, 1), respectively. Absolute reliability was evaluated using the Bland-Altman analysis. Statistical analyses were carried out using the R2.8.1 software. This research was approved by the ethics
committee of the Saitama Prefectural University (Approval number 24713), and all subjects provided written and oral informed consent.

RESULTS

The intra-measurement reliability, ICC (1, 1), for the E-E and T-T measurements was 0.972–0.992 and 0.900–0.977, respectively. The inter-measurement reliability, ICC (2, 1), for the E-E and T-T measurements was 0.914 and 0.725, respectively (Table 1). Overall, the E-E measurement had higher reliability than the T-T measurement. The validity study used the index point of the E-E measurement from the reliability study results. The intra-measurement reliability, ICC (1, 1), for the US and SE methods was 0.992 and 0.998, respectively. The inter-measurement reliability, ICC (2, 1), for both methods was 0.999 (Table 2). Blan-Altman analysis did not detect any proportional or fixed errors in the measurements. The 95% confidence interval for the minimal detectable change (MDC95) was 0.290 mm (Table 3).

DISCUSSION

This study assessed the reliability and validity of two methods with different index points using US-based imaging for quantitative evaluation of the interspinous process distance in the lumbar spine. The reliability study demonstrated no differences in the intra-rater reliability between the E-E and T-T measurements. However, the overall E-E measurement had a higher inter-rater reliability than the T-T measurement. This may be due to the shape of the lumbar spinous process. The top of the spinous process at this level was wide and round, which made defining it difficult. This affected the reliability of the interspinous process distance measurements. In addition, only 17 of the 30 images could be used for the T-T measurement despite the level selection based on the one that was most represented in the images. The remaining 13 images.

| Examiners | E-E measurement | T-T measurement |
|-----------|-----------------|-----------------|
| A         | 12.7 ± 3.6      | 29.6 ± 3.0      |
| B         | 11.7 ± 3.7      | 29.5 ± 2.3      |
| C         | 10.7 ± 3.7      | 29.9 ± 2.6      |
| D         | 11.6 ± 4.1      | 29.2 ± 2.8      |
| E         | 11.3 ± 3.7      | 30.1 ± 2.2      |

| n = 5     | E-E measurement | T-T measurement |
|-----------|-----------------|-----------------|
|           | 11.6 ± 3.7      | 29.6 ± 2.6      |

Table 1. Intra- and inter-rater reliability of two methods for measuring the lumbar interspinous process distance

| Measurement | Intra-measurement reliability |
|-------------|-------------------------------|
| US method   | 11.4 ± 4.0 0.992 [95% CI, 0.926–0.999] |
| SE method   | 11.4 ± 4.0 0.998 [95% CI, 0.979–1.000] |

| Measurement | Inter-measurement reliability |
|-------------|-------------------------------|
| n = 2       | 11.4 ± 4.0 0.999 [95% CI, 0.995–1.000] |

Table 2. Intra- and inter-measurement reliability of each method

| Fixed bias | Proportional bias | MDC95 (mm) |
|------------|-------------------|------------|
| −0.022–0.024 [no bias] | 0.004 (p = 0.915) [no bias] | 0.29 |

95% CI: 95% confidence interval; MDC95: the 95% confidence interval for the minimal detectable change

Table 3. Bland–Altman analysis of inter-measurement errors
which constituted 43% of the total images, were excluded at the stage of image selection. This approach has limited clinical applicability. However, the E-E measurement can be performed in the full lumbar flexion position. Therefore, the E-E measurement is the most suitable for lumbar spine mobility evaluations using US.

Next, this study assessed the validity of an US-based imaging method for quantitative assessment of interspinous process distances of the lumbar spine using the E-E measurement. It used both the US and SE methods to measure interspinous process distances in an MRLB. The validity of US images was assessed by comparing the results obtained using both methods.

MRLB differs from a real human body in two aspects. First, the bone material is different. Ultrasonic reflection is caused by a difference in the specific acoustic impedance; the reflected image of the bone surface is obtained because the specific acoustic impedance of the bone is lower than that of its surrounding tissues. In this study, the MRLB was made of a plastic skeletal model, paper towels, agar gel, and US gel, all with different specific acoustic impedances. Thus, the ultrasonic reflections obtained in this study can be considered to be similar to those from a real human bone. Therefore, the difference in the bone material did not significantly limit the validity of this study. The second difference was in the skin and subcutaneous tissues. In MRLB, the tissues were replaced by agar and US gels. Because the attenuation coefficients of the agar and US gels are lower than those of actual human tissues, it is possible that the images obtained from MRLB were clearer than those obtained from an actual human body. However, this issue does not limit the validity of the evaluation method under consideration.

A reliability coefficient of 0.41–0.60 is classified as moderate, 0.61–0.80 as substantial, and ≥0.81 as almost perfect. In this study, ICC (1, 1) was > 0.992 and ICC (2, 1) was 0.999. The ICCs, which reflected intra- and inter-examination reliability, were almost perfect, per the above-mentioned classification. In addition, absolute reliability was not detected by the Bland-Altman analysis. After assessing the random errors in measurement, an MDC95 value of 0.290 mm was obtained. Any changes smaller than this value cannot be reliably (confidence level, 95%) interpreted as actual changes in the score compared with random variations. The MDC95 value for the measurement accuracy of the US method was considerably lower (1 mm). MDC95 values for both methods were almost equivalent, indicating the validity of the US method.

Chleboun et al. reported a measurement method using a custom bed with a hole in its center. Accordingly, the measurement method was considered to have low clinical applicability. In contrast, the body position required for the examinations in this study did not require use of a custom bed with a hole in its center. In addition, the methods were found to have high reliability and validity. Therefore, the methods in this study had high clinical applicability.

The main limitation of the validity study was the inability to perform measurements in the lumbar extension and flexion positions. In the future, we will evaluate the validity of lumbar spine mobility evaluation methods in the actual human body. Additionally, the reliability study only measured the L2–3 segment. Therefore, the results cannot be generalized to other segments (L1–2, L3–4, and L4–5).

Our results demonstrate the high reliability and validity of US-based imaging methods in the quantitative assessment of interspinous process distances of the lumbar spine. In addition, the E-E measurement method was recommended. Therefore, this method can be used non-invasively to measure the interspinous process distance in vivo for the evaluation of lumbar spine mobility. Additionally, this measurement method may also be applicable during the initial evaluation of a LBP patient.

ACKNOWLEDGEMENT

This study was supported by the Society of Japanese Manual Physical Therapy.