J. Phys. Ther. Sci.
22: 301–306, 2010

Intrarater Reliability and Interrater Reliability in Spinal Motion Assessments

TAESUNG KO[1)], GUNSOO HAN[2)], BYUNGJUN CHO[3)], KWANWOO LEE[4)]

[1)] *Department of Physical Therapy, Daewon University College: 599 Shinwol-dong Jecheon-si Chungbuk 390-702, South Korea. TEL +82 43-649-3156*
[2)] *Department of Health Science, Kinesiology, Recreation and Dance, University of Arkansas*
[3)] *Department of Emergency Medical Technology, Kangwon National University*
[4)] *Department of Physical Therapy, Samyook University Graduate School*

Abstract. [Purpose] The aim of the present study was to measure the interrater and intrarater reliabilities of spinal motions using a skin-surface device (SpinalMouse; Idiag, Voletswil, Switzerland). [Methods] Spinal motion was measured in upright standing, trunk flexion, and trunk extension in 38 healthy adults. Intraclass correlation coefficients (ICCs) and Pearson's correlation coefficient were used to examine between-day and interrater reliabilities for spinal motion. [Results] For intrarater reliability of Raters 1 and 2, Pearson's correlation coefficient was 0.76–0.98, and ICC was 0.73–0.98, indicating high reliability in all cases except for extension of the sacral/hip. For interrater reliability on days 1 and 2, Pearson's correlation coefficient was 0.75–0.97 and ICC was 0.77–0.97 for all cases except for extension of the sacral/hip. [Conclusion] The reliability of SpinalMouse was demonstrated to be fair or higher but the measurement for extension of the sacral/hip needs further investigation.
Key words: Spinal motion, Skin-surface device, Reliability

(This article was submitted Feb. 15, 2010, and was accepted Mar. 24, 2010)

INTRODUCTION

Many methods exist for assessing spinal function. One of these methods is to measure the range of motion[1)]. Clinicians have routinely conducted visual analysis and radiographic assessments of spinal posture and back shape[2)]. Furthermore, many studies have been conducted on the range of motion of the spine[3–5)], but it is not easy to measure the range of motion of the spine and make a judgment about back shape because spinal motions are complex[6–7)].

Due to increasing concern about radiation exposure[2)], we need more objective and less invasive assessment methods than repetitive radiological measurements to monitor treatment effectiveness[8–10)]. Accordingly, devices to measure the treatment effectiveness of various spinal disorders through the skin surface are being developed[11)], as well as non-invasive techniques to measure spinal mobility, such as goniometer and inclinometer measurements to avoid the problem of repetitive radiation exposure[12–14)]. A variety of devices have been developed and used for spinal motion assessment, ranging from measurements using a simple tape to computer application devices[13,15–17)]. The advantages of these techniques are noninvasiveness, ease of use, and short assessment times[18)]. However, although most devices have been rated for reliability[13,19)], this reliability largely depends on the accuracy and the skin condition at the point where the device sensor is attached, and measurements may not be accurately performed[11)]. Moreover, kyphometers, goniometers, inclinometers and the like, cannot measure the entire spine, only a part of it such as the thoracic or lumbar spine[11].

A spinal motion analyzer (SpinalMouse; Idiag, Voletswil, Switzerland), It has been developed to solve these shortcomings, records the changes in length and inclination of the spine with a wheeled accelerometer, which moves along the spine[11]. It is a noninvasive device that can measure the mobility of the thoracic and lumbar spines at each segment in standing, flexion, extension, and lateral flexion[20]. This study was conducted to determine whether the SpinalMouse is an appropriate device for measuring spinal mobility from the thoracic to lumbar spines in clinical practice by analyzing the intrarater reliability and interrater reliability in the sagittal plane.

SUBJECTS AND METHODS

Subjects

This study selected 38 subjects from among 50 healthy adults aged 40 or younger, excluding 5 who had back pain, 1 who had experienced a fracture, and 6 who did not participate in either of the two measurements of each rater.

Methods

The SpinalMouse (Idiag, Voletswil, Switzerland) is a device connected to a computer through an analog-digital converter which measures the spinal shape along the paraspinal muscles in the standing position, flexion position, and extension position. This device was moved along the paraspinal muscles from C7 to S3 of the subjects to measure the angle of each segment along the outward form of the skin in the sagittal plane[20]. To test intrarater reliability and interrater reliability, measurements were performed on two consecutive days. For measurements in the standing position, the subjects were asked to stand upright with both feet open to shoulder width. For measurements in the flexion position, subjects were asked to slowly flex the trunk as much as possible, extend their knees with the neck slightly flexed while pointing their fingers toward their feet. For measurements in the extension position, the subjects were asked to slowly extended the trunk as much as possible, extend their knees with the neck slightly flexed without outside help. In each position, the SpinalMouse was positioned at C7 at the start and moved to S3 along the paraspinal muscles for measurement of the range of motion of each segment. SPSS 12.0 was used for statistical analysis. For the test-retest reliability, ICC (intra-class correlation coefficient) and Pearson's correlation coefficient were determined. For differences at each measurement point, the paired-t test was performed.

Table 1. General characteristics of subjects (n=38)

	Mean ± SD
Age (year)	30.29 ± 5.44
Height (cm)	170.47 ± 6.16
Weight (kg)	61.45 ± 11.08
BMI Index	22.77 ± 2.92
Male (n)	21
Female (n)	17

RESULTS

1. General characteristics of subjects

The subjects' mean age was 30.29, mean height 170.47 cm, mean weight 61.45 kg, and mean BMI index 22.77 (Table 1).

2. Intrarater reliability

For the intrarater reliability of Rater 1, Pearson's correlation coefficient was 0.76–0.88 for the thoracic spine, 0.73–0.90 for the lumbar spine, 0.67–0.91 for the sacral/hip, and 0.94–0.98 for the whole length. The ICC was 0.75–0.88 for the thoracic spine, 0.73–0.88 for the lumbar spine, 0.66–0.84 for the sacral/hip, and 0.94–0.96 for the whole length. The difference in mean values between the measurement times was significant only for the sacral/hip in the standing position, and other differences were insignificant (Table 2).

For the intrarater reliability of Rater 2, Pearson's correlation coefficient was 0.77–0.92 for the thoracic spine, 0.84–0.92 for the lumbar spine, 0.60–0.84 for the sacral/hip, and 0.94–0.97 for the whole length. The ICC was 0.77–0.92 for the thoracic spine, 0.83–0.92 for the lumbar spine, 0.59–0.84 for the sacral/hip, and 0.94–0.97 for the whole length. The differences in mean values between the measurement times were insignificant (Table 3).

3. Interrater reliability

For the interrater reliability on day 1, Pearson's correlation coefficient was 0.79–0.90 for the

Table 2. Intrarater reliability of Rater 1

Rater 1	Day 1	Day 2	Pearson's *r*	ICC (95% CI)
T-Spine				
Standing	35.29 ± 9.49	34.68 ± 8.85	0.77**	0.77 (0.60 – 0.87)
Flexion	59.68 ± 9.93	60.66 ± 8.42	0.76**	0.75 (0.57 – 0.86)
Extension	39.53 ± 16.86	39.68 ± 16.22	0.88**	0.88 (0.79 – 0.93)
L-Spine				
Standing	–9.76 ± 13.20	–11.03 ± 11.09	0.87**	0.86 (0.75 – 0.92)
Flexion	42.47 ± 12.92	40.61 ± 11.66	0.73**	0.73 (0.53 – 0.85)
Extension	–27.86 ± 17.81	–27.92 ± 14.99	0.90**	0.88 (0.79 – 0.94)
Sacral/hip				
Standing	2.24 ± 10.21	3.82 ± 9.02*	0.91**	0.91 (0.83 – 0.95)
Flexion	38.63 ± 16.90	37.95 ± 17.65	0.75**	0.75 (0.56 – 0.82)
Extension	–4.82 ± 8.99	–4.84 ± 7.33	0.67**	0.66 (0.43 – 0.80)
Whole Length				
Standing	524.00 ± 72.06	524.53 ± 70.50	0.98**	0.98 (0.96 – 0.99)
Flexion	620.21 ± 92.35	617.95 ± 82.65	0.94**	0.94 (0.88 – 0.96)
Extension	511.71 ± 76.88	508.76 ± 69.32	0.97**	0.96 (0.93 – 0.98)

Values are degrees. * p<0.05. ** p<0.01.

Table 3. Intrarater reliability of Rater 2

Rater 2	Day 1	Day 2	Pearson's *r*	ICC(95% CI)
T-Spine				
Standing	34.13 ± 11.03	34.81 ± 10.05	0.77**	0.77 (0.60 – 0.87)
Flexion	60.68 ± 8.59	60.57 ± 8.97	0.77**	0.77 (0.61 – 0.87)
Extension	40.18 ± 18.04	40.13 ± 17.20	0.92**	0.92 (0.85 – 0.95)
L-Spine				
Standing	–12.13 ± 9.88	–13.21 ± 11.17	0.84**	0.83 (0.71 – 0.91)
Flexion	40.47 ± 11.20	39.45 ± 11.22	0.85**	0.85 (0.73 – 0.92)
Extension	–27.36 ± 14.51	–28.58 ± 14.25	0.92**	0.92 (0.85 – 0.95)
Sacral/hip				
Standing	4.26 ± 8.79	5.45 ± 9.04	0.84**	0.84 (0.71 – 0.91)
Flexion	40.50 ± 16.43	38.37 ± 16.36	0.88**	0.88 (0.79 – 0.94)
Extension	–8.08 ± 6.63	–6.05 ± 7.76	0.60**	0.59 (0.34 – 0.76)
Whole Length				
Standing	524.82 ± 70.42	521.55 ± 70.76	0.97**	0.97 (0.94 – 0.98)
Flexion	617.05 ± 76.56	613.89 ± 78.36	0.95**	0.95 (0.92 – 0.97)
Extension	507.47 ± 66.50	503.63 ± 67.24	0.94**	0.94 (0.90 – 0.97)

Values are degrees. ** p<0.01.

thoracic spine, 0.81–0.84 for the lumbar spine, 0.58–0.83 for the sacral/hip, and 0.96–0.97 for the whole length. The ICC was 0.78–0.90 for the thoracic spine, 0.80–0.81 for the lumbar spine, 0.55–0.82 for the sacral/hip, and 0.94–0.97 for the whole length. The differences in mean values between raters were significant for the lumbar spine in the standing position and for the sacral/hip in the standing and extension positions; other differences were insignificant (Table 4).

For the interrater reliability on day 2, Pearson's correlation coefficient was 0.84–0.93 for the thoracic spine, 0.77–0.93 for the lumbar spine, 0.88–0.90 for the sacral/hip, and 0.95–0.97 for the whole length. The ICC was 0.83–0.92 for the thoracic spine, 0.77–0.92 for the lumbar spine,

Table 4. Interrater reliability of Day 1

Day 1	Rater 1	Rater 2	Pearson's r	ICC(95% CI)
T-Spine				
Standing	35.29 ± 9.49	34.13 ± 11.03	0.79**	0.78 (0.62 – 0.88)
Flexion	59.68 ± 9.93	60.68 ± 8.59	0.79**	0.78 (0.62 – 0.88)
Extension	39.52 ± 16.86	39.53 ± 16.86	0.90**	0.90 (0.82 – 0.94)
L-Spine				
Standing	–9.76 ± 13.20	–12.13 ± 9.89*	0.84**	0.81 (0.66 – 0.89)
Flexion	42.47 ± 12.92	40.47 ± 11.20	0.81**	0.80 (0.65 – 0.89)
Extension	–27.87 ± 17.81	–27.37 ± 14.51	0.83**	0.81 (0.67 – 0.90)
Sacral/hip				
Standing	2.24 ± 10.21	4.26 ± 8.79*	0.83**	0.82 (0.68 – 0.90)
Flexion	38.63 ± 16.90	40.50 ± 16.43	0.75**	0.75 (0.57 – 0.86)
Extension	–4.82 ± 8.99	–8.08 ± 6.63*	0.58**	0.55 (0.29 – 0.74)
Whole Length				
Standing	524.00 ± 72.06	524.82 ± 70.42	0.97**	0.97 (0.95 – 0.98)
Flexion	620.21 ± 92.35	617.50 ± 76.56	0.96**	0.94 (0.90 – 0.97)
Extension	511.71 ± 76.88	507.47 ± 66.50	0.96**	0.95 (0.91 – 0.97)

Values are degrees. * $p<0.05$. ** $p<0.01$.

Table 5. Interrater reliability of Day 2

Day 2	Rater 1	Rater 2	Pearson's r	ICC (95% CI)
T-Spine				
Standing	34.68 ± 8.85	34.82 ± 10.05	0.85**	0.84 (0.72 – 0.91)
Flexion	60.66 ± 8.42	60.58 ± 8.97	0.84**	0.83 (0.71 – 0.91)
Extension	39.68 ± 16.22	40.13 ± 17.20	0.93**	0.92 (0.86 – 0.96)
L-Spine				
Standing	–11.03 ± 11.09	–13.21 ± 11.17*	0.85**	0.85 (0.74 – 0.92)
Flexion	40.61 ± 11.66	39.45 ± 11.22	0.77**	0.77 (0.60 – 0.87)
Extension	–27.92 ± 14.99	–28.57 ± 14.25	0.93**	0.92 (0.86 – 0.96)
Sacral/hip				
Standing	3.82 ± 9.02	5.45 ± 9.04*	0.88**	0.88 (0.79 – 0.94)
Flexion	37.95 ± 17.65	38.37 ± 16.36	0.90**	0.90 (0.81 – 0.94)
Extension	–4.84 ± 7.33	–6.05 ± 7.76*	0.90**	0.89 (0.81 – 0.94)
Whole Length				
Standing	524.53 ± 70.50	521.55 ± 70.76	0.97**	0.97 (0.94 – 0.98)
Flexion	617.95 ± 82.65	613.90 ± 78.36	0.95**	0.95 (0.91 – 0.97)
Extension	508.76 ± 69.32	503.63 ± 67.24	0.96**	0.96 (0.93 – 0.98)

Values are expressed degrees. * $p<0.05$. ** $p<0.01$.

0.88–0.90 for the sacral/hip, and 0.95–0.97 for the whole length. The differences in mean values between raters were significant for the lumbar spine in the standing position and for the sacral/hip in the standing and extension positions; other differences were insignificant (Table 5).

DISCUSSION

This study was conducted to determine the intrarater reliability and interrater reliability of the SpinalMouse, which assesses spinal mobility in a noninvasive manner. For intrarater reliability of Raters 1 and 2, Pearson's correlation coefficient was 0.76 or higher, and ICC was 0.73 or higher, indicating high reliability in all cases except for extension of the sacral/hip (Rater 1: ICC=0.66, r= 0.67, Rater 2: ICC= 0.59, r=0.60). For interrater reliability on days 1 and 2, Pearson's correlation coefficient was 0.75 or higher and ICC was 0.77 or higher for all cases except for extension of the sacral/hip (Day 1: ICC=0.55, r=0.58, Day 2: ICC=0.89, r=0.90). In previous studies, Currier (1990)[21] stated that 0.90–0.99 = high reliability, 0.80–0.89 = good reliability, 0.70–0.79 = fair reliability, and <0.69 = bad reliability. Chinn (1991)[22] reported that a device was useful if the ICC was > 0.6. Based on these studies, we can say that the SpinalMouse is reliable and useful for clinical applications.

In previous studies of devices for measuring the range of spinal motion, Keeley et al.(1986)[23] reported r=0.92 (p<0.001) for inter-rater correlation using an inclinometer, which is higher than the correlation found in this study. In a study using a single-sensor inclinometer, Chen et al.(1997)[12] reported ICC=0.39 for extension and ICC=0.69 for bending in terms of interrater reliability, which are lower than the interrater reliabilities found in this study. Furthermore, Dopf et al.(1994)[24] reported ICC=0.84 for extension and ICC=0.76 for bending in terms of interrater reliability using a spine motion analyzer, which are lower than the interrater reliabilities found in this study. These previous studies show that the reliability of various measurement devices used for the same purpose can vary depending on the characteristics of the device and the measurement methods. Therefore, before using any measurement device in the clinical setting, we need to measure reliability in various methods, positions, and situations, and then choose an appropriate device for the purpose.

Mannion et al.(2004)[11] using a SpinalMouse measured 20 healthy people and reported that the ICC of the first (day 1) and second (day 2) measurements by Rater 1 ranged from 0.67 to 0.92 (mean: 0.82). The ICC of Rater 2 ranged from 0.57 to 0.95 (mean: 0.83). Further, the ICCs of Rater 1 and Rater 2 ranged from 0.62 to 0.92 (mean: 0.81) for the first measurements (day 1) and from 0.70 to 0.94 (mean: 0.86) for the second measurement (day 2). These results are similar to those of the current study. Furthermore, in a study using the SpinalMouse, Post et al.(2004)[20] reported ICC=0.95, r=0.90 (p<0.001) for bending, ICC=0.92, r=0.85(p<0.001) for extension, and ICC=0.76, r=0.61(p<0.001) for the whole length, which are higher than the findings in this study. However, since they did not differentiate between thoracic and lumbar parts of the spine, therefore a direct comparison with this study cannot be made.

The reason for the lower reliability in the case of sacral/hip extension can be found in previous studies, which show that to maintain an upright standing position by extending the trunk requires a balance ability, and it is difficult to maintain such balance[20]. Further, trunk extension in an upright standing position is inconvenient and it is difficult to maintain balance[18,25]. In other words, the subjects' understanding of, and ability to repeat the measurement positions, is an important factor that must be sufficiently explained to the subjects before measurement.

Post et al.(2004)[20] claimed that we need minute palpation of the start and end points of a measurement to achieve an accurate measurement, and that the movement path may change when the measurement device is moved along the spine. In addition, Mayer et al.(1995)[26] stated that the diversity of the starting points of measurement among measurers is also a major influencing factor when externally measuring the range of motion of the spine. These studies indicate that raters need sufficient knowledge of the measurement methods and anatomy as well as knowledge about the individual subjects.

Based on the finding that the repeatability of the same position during two consecutive measurements is a major factor that influences the measurement of spinal mobility[27] and the finding that subjects cannot accurately repeat the same position each time[20], the raters must have an understanding of the measurement device and basic knowledge of the human body to achieve accurate measurement in consideration of the body characteristics of the subjects. Reliability will improve if subjects can repeat the same position each time. However, if some positions are difficult for the subjects to adopt accurately, reliability can decrease.

This study found that the SpinalMouse is a highly reliable measurement device for spinal mobility and is useful in clinical settings. Further studies will be necessary to measure the reliability of the device for other diseases.

REFERENCES

1) Gill KP, Callaghan MJ: Intratester and intertester reproducibility of the lumbar motion monitor as a measure of range, velocity and acceleration of the thoracolumbar spine. Clin Biomech (Bristol, Avon), 1996, 11(7): 418–421.
2) Morin Doody M, Lonstein JE, Stovall M, et al.: Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine, 2000, 25(16): 2052–2063.
3) Knop C, Fabian HF, Bastian L, et al.: Late results of thoracolumbar fractures after posterior instrumentation and transpedicular bone grafting. Spine, 2001, 26(1): 88–99.
4) Leferink VJ, Zimmerman KW, Veldhuis EF, et al.: Thoracolumbar spinal fractures: radiological results of transpedicular fixation combined with transpedicular cancellous bone graft and posterior fusion in 183 patients. Eur Spine J, 2001, 10(6): 517–523.
5) Lim TH, Eck JC, An HS, et al.: A noninvasive, three-dimensional spinal motion analysis method. Spine, 1997, 22(17): 1996–2000.
6) Salisbury PJ, Porter RW: Measurement of lumbar sagittal mobility. A comparison of methods. Spine, 1987, 12(2): 190–193.
7) Stokes IA, Bevins TM, Lunn RA: Back surface curvature and measurement of lumbar spinal motion. Spine, 1987, 12(4): 355–361.
8) Goldberg CJ, Kaliszer M, Moore DP, et al.: Surface topography, cobb angles, and cosmetic change in scoliosis. Spine, 2001, 26: E55–E63.
9) Goldberg MS, Mayo NE, Poitras B: The Ste-Justine adolescent idiopathic scoliosis cohort study. II. Perception of health, self and body image, and participation in physical activities. Spine, 1994, 19(14): 1562–1572.
10) Pratt RK, Burwell RG, Cole AA: Patient and parental perception of adolescent idiopathic scoliosis before and after surgery in comparison with surface and radiographic measurements. Spine, 2002, 27(14): 1543–1552.
11) Mannion AF, Knecht K, Balaban G, et al.: A new skin-surface device for measuring the curvature and global and segmental ranges of motion of the spine: reliability of measurements and comparison with data reviewed from the literature. Eur Spine J, 2004, 13(2): 122–136.
12) Chen SP, Samo DG, Chen EH, et al.: Reliability of three lumbar sagittal motion measurement methods: surface inclinometers. J Occup Environ Med, 1997, 39(3): 217–223.
13) Mannion A, Troke M: A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. Clin Biomech , 1999, 14(9): 612–619.
14) Mayer TG, Kondraske G, Beals SB, et al.: Spinal range of motion. Accuracy and sources of error with inclinometric measurement. Spine, 1997, 22(17): 1976–1984.
15) Schuit D, Petersen C, Johnson R, et al.: Validity and reliability of measures obtained from the OSI CA-6000 Spine Motion Analyzer for lumbar spinal motion. Man Ther, 1997, 2(4): 206–215.
16) Troke M, Moore AP, Maillardet FJ, et al.: A new, comprehensive normative database of lumbar spine ranges of motion. Clin Rehabil, 2001, 15(4): 371–379.
17) Van Herp G, Rowe P, Salter P, et al.: Three-dimensional lumbar spinal kinematics: a study of range of movement in 100 healthy subjects aged 20 to 60+ years. Rheumatology (Oxford), 2000, 39(12): 1337–1340.
18) Ng JK, Kippers V, Richardson CA, et al.: Range of motion and lordosis of the lumbar spine: reliability of measurement and normative values. Spine, 2001, 26(1): 53–60.
19) Essendrop M, Maul I, Läubli T, et al.: Measures of low back function: a review of reproducibility studies. Clin Biomech, 2002, 17(4): 235–249.
20) Post RB, Leferink VJ: Spinal mobility: sagittal range of motion measured with the SpinalMouse, a new non-invasive device. Arch Orthop Trauma Surg, 2004, 124(3): 187–192.
21) Currier DP: Elements of research in physical therapy. Third ed. Baltimore: Williams and Wilkins, 1990, pp167.
22) Chinn S: Repeatability and method comparison. Thorax, 1991, 46(6): 454–456.
23) Keeley J, Mayer TG, Cox R, et al.: Quantification of lumbar function. Part 5: Reliability of range-of-motion measures in the sagittal plane and an in vivo torso rotation measurement technique. Spine, 1986, 11(1): 31–35.
24) Dopf CA, Mandel SS, Geiger DF, et al.: Analysis of spine motion variability using a computerized goniometer compared to physical examination. A prospective clinical study. Spine, 1994, 19(5): 586–595.
25) Dillard J, Trafimow J, Andersson GB, et al.: Motion of the lumbar spine. Reliability of two measurement techniques. Spine, 1991, 16(3): 321–324.
26) Mayer RS, Chen IH, Lavender SA, et al.: Variance in the measurement of sagittal lumbar spine range of motion among examiners, subjects, and instruments. Spine, 1995, 20(13): 1489–1493.