

Original Article

Investigation of absolute intra-rater and inter-rater reliabilities during the muscle hardness estimation

KAZUNORI MOROZUMI, RPT, PhD^{1)*}, KATSUYUKI MORISHITA, RPT, PhD²⁾,
MASAHIRO AOKI, RPT, MS³⁾, YUKA YOKOI, RPT, PhD²⁾, TAKASHI INOMATA, RPT, PhD¹⁾

¹⁾ Department of Physical Therapy, School of Health Sciences, Tokyo International University: 1-13-1 Matobakita, Kawagoe, Saitama 350-1197, Japan

²⁾ Department of Physical Therapy, Faculty of Social Work Studies, Josai International University, Japan

³⁾ Department of Rehabilitation, Hachioji Health Cooperative Shiroyama Hospital, Japan

Abstract. [Purpose] This study aimed to investigate the absolute intra-rater and inter-rater reliabilities during the measurement of muscle hardness, which is used to evaluate physical therapy. Moreover, we examined the effects of using different equipment types and their positioning on the intra-rater and inter-rater reliabilities. [Participants and Methods] Participants of this study comprised 12 healthy adult male individuals. Two experts and two beginners measured the muscle hardness of the lumbar erector spinae and rectus femoris using three types of hardness meters at two positions, including when the muscle was relaxed and stretched. [Results] Intra-rater fixed bias was observed during some measurements by both experts and beginners. Inter-rater fixed bias was observed during measurements by some experts and not the beginners. [Conclusion] In this study, the measurement of muscle hardness demonstrated a need to reconsider the measurement position and acclimation time. These examinations require the consideration of relative and absolute reliabilities.

Key words: Muscle hardness measurement, Tissue hardness meter, Absolute reliability

(This article was submitted Sep. 16, 2021, and was accepted Nov. 13, 2021)

INTRODUCTION

Currently, the tissue hardness measurements and pain evaluations, which are used to assess the effectiveness of ongoing physiotherapy, may rely on subjective evaluations of physiotherapy, including palpation by the rater and complaints. Such methods are widely used because they can be conducted without needing measurement equipment. However, adequate knowledge, experience, and techniques are needed to perform highly reproducible and accurate assessments. Accordingly, objective measurements using a tissue hardness meter or an algometer have been considered^{1–17)}. Among the procedures that measure muscle hardness is ultrasound elastography, which quantitatively forms images of tissue hardness using ultrasonography and magnetic resonance imaging¹⁸⁾. This method records the biological response caused by a quick mechanical impact with a weak force from the outside as a damped-free vibration of the live tissue. Moreover, it uses a so-called push-in type tissue hardness meter^{11, 19, 20)} that measures pressure when pushed in, assuming that it is a linear elastic body¹⁹⁾ that has an elastic and viscous response to soft tissues, such as muscles.

The methods for measuring muscle and soft tissue hardness using a push-in tissue hardness meter has been previously examined by our team^{10, 16, 17, 21, 22)}. In these studies, we proposed the use of proper measurement equipment to ensure highly reliable and reproducible measurements¹⁰⁾. Limb position and measurement procedures¹⁶⁾, including the validity of

*Corresponding author. Kazunori Morozumi (E-mail: moroz@nifty.com)

©2022 The Society of Physical Therapy Science. Published by IPEC Inc.



This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (by-nc-nd) License. (CC-BY-NC-ND 4.0: <https://creativecommons.org/licenses/by-nc-nd/4.0/>)

pre-measurement acclimation time and quantitative evaluation, among others^{17, 21, 22}), should be considered during measurements. Producing highly reliable measurements using push-in tissue hardness meters may require technical proficiency. In our previous work, we reported the effects of proficiency in muscle hardness measurement, different instruments, and positioning on the intra-rater and the inter-rater reliability and revealed that differences in proficiency may affect relative reliability²³. Particularly, the relative intra- and inter-rater reliabilities in the beginner group exhibited high reliability compared with the average of three measurements, suggesting the importance of increasing the average number of measurements and also indicating the appropriate number of measurements. By assessing the changes in the muscle tone while the limb position during measurement (i.e., when the muscle is relaxed or stretched) was changing, it was shown that the relative reliability was better in the expert and beginner groups when the muscle was stretched than when it was relaxed.

In this study, we used three types of tissue hardness meters operating based on different measurement methods to examine the data of beginners having no lectures or practical experience regarding muscle hardness measurements and those of experts who had experience in measuring muscle hardness using such equipment. We aimed to ascertain the issues that should be addressed to realize highly objective muscle hardness measurements by performing measurements using each type of equipment while changing the limb position during measurement. Then, we examined the absolute reliability of these measurements.

PARTICIPANTS AND METHODS

Male individuals with a body mass index (BMI) of ≥ 18.5 and < 25 kg/m², aged 20–39 years, and without a history of trauma or disease with residual movement disorders in the four limbs, were included. We used these criteria to minimize any errors resulting from differences in physical composition between men and women and age distribution. The participants were 12 healthy adult male individuals with a mean age, height, body weight, and BMI of 20.0 ± 0.0 years, 172.9 ± 6.2 cm, 65.0 ± 8.1 kg, and 21.7 ± 2.2 kg/m², respectively. The raters were two experts (expert group) having 16–18 years of teaching and research experience using tissue hardness meters as well as two beginners (beginner group) in the 2nd year of their studies in the Department of Physiotherapy who applied to participate in the study and had no lectures or practical experience related to muscle hardness measurement. Before the experiment, each rater had the opportunity to be trained on how to use each device and how to take measurements.

Prior to study initiation, the participants were informed concerning the purpose and details of this study, including its benefits and risks; we also informed them concerning the protection of personal information and the option to refuse to participate or withdraw from the study. Then, they provided informed consent for participation. This study was approved by the research ethics committee of the Josai International University (approval number: 01P180017).

Three types of tissue hardness meters with different measurement modes (Equipment A: Myoton-Pro, Myoton AS, Tallinn, Estonia; Equipment B: PEK-MP, Imoto Seisakusho Corp., Tokyo, Japan; and Equipment C: NEUTONE TDM-Z2, TRY-ALL, Chiba, Japan) were used. Equipment A is a tissue hardness meter that records the biological response caused by a quick mechanical impact (tap time, 15 ms) as a damped-free vibration of live tissue, and calculates the parameters related to tissue viscoelasticity from the shape of the acceleration signal. Equipment B and Equipment C are push-in type tissue hardness meters. With these meters, measurements could be performed automatically simply by turning the handle slowly to minimize errors because of differences in pushing force and tester speed while performing the measurements.

In the lower limbs and body trunk, we took measurements from the two test muscles of the right rectus femoris and right lumbar erector spinae, respectively, considering the resting limb position and ability to expose the measurement surface. The measurement point of the right lumbar erector spinae was 3 cm to the right lateral side of the spinous process on Jacoby's line. The measurement point of the right rectus femoris was the midpoint on the line connecting the anterior inferior iliac spine and the center point of the suprapatellar margin. Two muscle hardness measurements were obtained from each of these points. Figure 1 shows the position of the limb during measurement. The following two limb positions were used: the relaxed position (R position), in which each test muscle is in its anatomical basic limb position and in a state of low muscle tension, and the stretched limb position (S position), in which the muscle being measured is stretched, the muscle tension is elevated, and measurement reliability is thought to increase^{15, 16}. We measured the right lumbar erector spinae with participants in the prone position. In the R position, the leg trunk flexion/extension rotation intermediate position and the hip and knee joint flexion/extension position were at 0°, with the ankle joint in a position allowing the foot to hang from the edge of the bed. The R position was set so as to ensure that the participants rotated the neck leftwards to put the check on the bed and placed both upper limbs to the side of the body. To measure the right lumbar erector spinae in the S position, the participants had a semi-conical stretch pole (bottom square dimensions: length, 40 cm; width, 15 cm; cone radius: height, 7.5 cm) inserted to the abdominal iliac crest when lying in the aforementioned prone position, flexing the lumbar spine and hip joint and stretching the lumbar erector spinae. The R position of the right rectus femoris was achieved by having participants laid in a resting supine position on a flat mat. The S position was considered the same as the Thomas position, in which the participant himself, with his lower leg below both knee joints hanging down from the edge of the bed, uses both arms to hold the anterior surface of the left lower leg, which is the side opposite to the measurement side, to flex the hip and knee joints; it is a limb position that changes the anterior tilt of the pelvis to a horizontal tilt and stretches the right rectus femoris.

For each measurement, a randomization chart was used to randomize the combination of the rater, measurement equip-

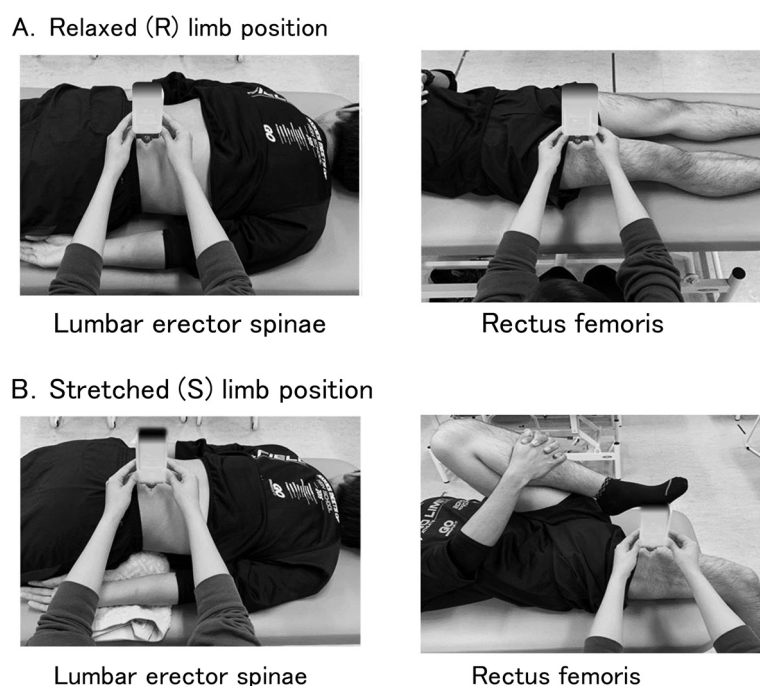


Fig. 1. Relaxed (R) limb position and stretched (S) limb position during muscle hardness measurement in each muscle.

ment, and measurement order. Prior to initiating measurements, the Principal Investigator marked all measurement points in each participant, and the participants were provided 5 min of acclimation time in a resting supine position. Especially, one rater took measurements from one participant by assessing all test muscles in both limb positions with the same tissue hardness meter. Subsequently, the participants were provided 5 min of acclimation time for rest, and measurements were repeated with changes in measurement equipment. For equipment other than Equipment A, hardness (N/m) was measured and calculated by placing the results of the measurement in the formulas of the respective approximate curves: Equipment B: $y=(4.0405x+93.373)\times 100$; Equipment C: $y=(0.0238x+0.532)\times 100$.

For the absolute intra-rater and inter-rater reliabilities, systematic bias (fixed bias and proportional bias) and minimal detectable change 95 (MDC₉₅)^{24, 25} were calculated according to the method proposed by Shimoi et al.^{24, 26}. The latter comprises the use of the Bland–Altman analysis on differences in reliability because of differences in limb position during measurement and measuring equipment in the beginner and expert groups. A Bland–Altman analysis of absolute intra-rater reliability was conducted for two items between the first and second measurements by a given rater. The Bland–Altman analysis of absolute inter-rater reliability was performed by Beginners I and II as well as by Experts I and II by comparing the arithmetic mean outcome of the two measurements performed by each rater. The Bland–Altman analysis was used to calculate two items, the average (x-axis) and difference (y-axis) of the two measurements taken by each rater and that of two different raters. We created a scatter plot by plotting the data of these two items in each group according to each measurement equipment and test muscle. To examine whether there was fixed bias, we determined the 95% confidence interval of the average of differences between both items. In cases where this difference was not 0, we considered the measured values to be distributed in a certain direction (i.e., positively or negatively from the x-axis where the measurements match exactly); accordingly, we deemed that there was fixed bias^{24, 26}. To determine whether there was proportional bias, we calculated the regression formula of the Bland–Altman plot created, tested the significance of the regression, and determined that significant regression indicated the presence of proportional bias^{24, 26}. We compared the muscle hardness in the R and S positions between each equipment and tested muscle based on absolute inter-rater reliability using MDC₉₅. We calculated differences in muscle hardness between the R and S positions and determined that differences that exceed the MDC₉₅ of the R position measurement indicated a true change, not a measurement error^{24, 25}.

RESULTS

Table 1 shows the results of the absolute intra-rater reliability (presence or absence of fixed bias and proportional bias), as determined by the Bland–Altman analysis of measurements taken by the expert group. Fixed bias, which is an indicator of absolute intra-rater reliability, was identified in three measurements taken by Expert I (i.e., in the measurement of the lumbar erector spinae with Equipment A and in the measurements of the lumbar erector spinae and rectus femoris with Equipment C,

all in the R position). In the S position, fixed bias was observed in two measurements of the rectus femoris using Equipment B and C, respectively. Fixed bias was observed in the measurement taken by Expert II of the rectus femoris in R position using Equipment C. Proportional bias, as determined by intra-rater reliability, was not observed in any of the items measured in the expert group.

Table 2 presents the results of absolute intra-rater reliability (presence or absence of fixed bias and proportional bias) as assessed by the Bland–Altman analysis of the measurements taken by the beginner group. Fixed bias was noted only in one measurement taken by Beginner I (i.e., the measurement of the lumbar erector spinae in the S position using Equipment B). Concerning Beginner II, fixed bias was observed in two measurements (i.e., the measurement of the lumbar erector spinae in the R position using Equipment A and that made in the rectus femoris in the R position using Equipment C; nevertheless, there was no fixed bias for this rater in any tested muscle in the S position. Proportional bias in intra-rater reliability was not observed for any item in the beginner group.

Table 3 and Figs. 2–4 show the results of the absolute inter-rater reliability (presence or absence of fixed bias and proportional bias) as assessed by the Bland–Altman analysis of the measurements taken by the expert group and beginner group. Fixed bias in the expert group, which is an indicator of absolute inter-rater reliability, was identified in two measurements (i.e., the measurement of the rectus femoris in the R position using Equipment B and C). In the S position, fixed bias was observed in three measurements (i.e., the measurement of the lumbar erector spinae and rectus femoris using Equipment B and that of the rectus femoris using Equipment C). Proportional and fixed biases in inter-rater reliability were not observed in any item in the expert and beginner groups, respectively. In the beginner group, proportional bias was observed only in one measurement (i.e., the measurement of the rectus femoris in the S position using Equipment B).

Table 4 shows the differences calculated by subtracting the measurement in the R position from that in the S position and the MDC₉₅ values in the expert and beginner groups. For both the expert and beginner groups, the values obtained by the aforementioned subtraction exceeded the MDC₉₅ values of all measurement equipment.

DISCUSSION

Shimoi²⁴⁾ explains that the measured values comprise the true value along with errors. Errors are roughly divided into random errors and systematic biases. Especially, random errors are fluctuations and variations that occur relatively symmetrically in both directions from the average value and are further divided into biological individual differences and measurement errors. Systematic biases are structural and systematic deviations from the true value and can be divided into fixed and proportional biases. It has been reported that fixed bias is an error that may occur in a specific direction regardless of the magnitude of the true value and that proportional bias is an error that increases proportionally with the magnitude of the true value²⁴⁾.

In this study, proportional bias in intra-rater reliability was not observed for any item in the beginner and expert groups, suggesting that there was no error proportionally increasing with the magnitude of the true value. In contrast, in both groups, some measurements had fixed bias in intra-rater reliability. It is possible that these fixed biases were caused by changes in muscle tension because of differences in limb positions during measurement. Measurements taken at low muscle tension are more susceptible to the effects of bodily structures other than muscles, such as bones and fat, and it is possible that the values may be biased in a specific direction depending on these situations. In the expert group, fixed bias was observed in four measurements in the R position; however, fixed bias was observed for only two measurements in the S position. In the beginner group, fixed bias was observed in two measurements in the S position, yet this was only true for one measurement in the S position. In the expert and beginner groups, fixed bias was less in the R than in the S position, suggesting that increased muscle tension might have caused fixed bias. Particularly, there was no fixed bias in any of the measurements taken in the S position using Equipment A, indicating that fixed bias in intra-rater reliability can be improved by taking measurements in the S position and that using Equipment A can help improve fixed bias even further. However, although biological surface hardness is highly reliable and reproducible in the measurements taken by Equipment A, there are problems associated with its use as evidenced by reports. Especially, it may be difficult to detect hardness in the vertical direction as the thickness and hardness of the object being measured increase²⁷⁾. Moreover, Equipment A is more expensive compared with Equipment B and Equipment C.

In the expert group, fixed bias, which is an indicator of absolute inter-rater reliability, was observed in two and three measurements taken in the R and S positions, respectively. In contrast, fixed bias in inter-rater reliability was not observed in any of the measurements taken by the beginner group. Based on Figs. 2–4, by showing the visual observations related to these outcomes, we can discern that many of the plot measurements taken by the expert group that had fixed biases were distributed in the negative direction. However, in the beginner group, distribution was observed between both positive and negative directions. A simple interpretation of these observations shows that the beginner group could perform accurate measurements with less fixed and systematic biases than the expert group. We have previously reported²³⁾ that the relative reliability of measurements in the beginner group tended to decrease because of differences in proficiency between the groups. Accordingly, the hypothesis that this trend is different from the fixed bias in the inter-rater reliability found in this study may be misleading. As beginners are unfamiliar with the handling of equipment during measurement, changing the measurement interval and limb position needed a considerable amount of time. We conceived that such time-exhaustive measurements

Table 1. Absolute intra-rater reliability of measurements taken by Expert I and Expert II (presence or absence of fixed bias and proportional bias), as determined by the Bland–Altman analysis

	Relaxed (R) limb position						Stretched (S) limb position					
Expert I												
	Fixed bias			Proportional bias			Fixed bias			Proportional bias		
(Equipment A)	95% confidence interval		t value	p value		95% confidence interval		t value	p value			
Lumbar erector spinae	-16.125 to -3.208	presence	-1.713	p≥0.05	absence	-9.39 to 14.56	absence	-0.488	p≥0.05	absence		
Rectus femoris	-3.395 to 3.229	absence	-1.888	p≥0.05	absence	-1.79 to 30.46	absence	-0.249	p≥0.05	absence		
(Equipment B)												
Lumbar erector spinae	-0.149 to 0.546	absence	-0.174	p≥0.05	absence	-0.404 to 0.206	absence	1.236	p≥0.05	absence		
Rectus femoris	-0.479 to 0.347	absence	0.466	p≥0.05	absence	-0.692 to -0.033	presence	0.456	p≥0.05	absence		
(Equipment C)												
Lumbar erector spinae	0.338 to 6.008	presence	0.340	p≥0.05	absence	0.131 to 5.422	presence	1.352	p≥0.05	absence		
Rectus femoris	0.395 to 4.760	presence	-0.297	p≥0.05	absence	-2.073 to 3.263	absence	1.193	p≥0.05	absence		
Expert II												
	Fixed bias			Proportional bias			Fixed bias			Proportional bias		
(Equipment A)	95% confidence interval		t value	p value		95% confidence interval		t value	p value			
Lumbar erector spinae	-7.399 to 9.399	absence	0.902	p≥0.05	absence	-5.483 to 14.31	absence	1.671	p≥0.05	absence		
Rectus femoris	-2.387 to 2.720	absence	-1.285	p≥0.05	absence	-8.338 to 2.50	absence	-0.688	p≥0.05	absence		
(Equipment B)												
Lumbar erector spinae	-0.793 to 0.2655	absence	-0.651	p≥0.05	absence	-0.244 to 0.508	absence	-1.613	p≥0.05	absence		
Rectus femoris	-0.590 to 0.458	absence	-0.116	p≥0.05	absence	-0.402 to 0.468	absence	1.037	p≥0.05	absence		
(Equipment C)												
Lumbar erector spinae	-0.486 to 3.263	absence	-1.064	p≥0.05	absence	-2.380 to 1.984	absence	0.886	p≥0.05	absence		
Rectus femoris	0.466 to 5.086	presence	-0.189	p≥0.05	absence	-2.336 to 2.733	absence	1.42	p≥0.05	absence		

Table 2. Absolute intra-rater reliability of measurements taken by Beginner I and Beginner II (presence or absence of fixed bias and proportional bias), as determined by the Bland–Altman analysis

	Relaxed (R) limb position						Stretched (S) limb position					
Beginner I												
	Fixed bias			Proportional bias			Fixed bias			Proportional bias		
(Equipment A)	95% confidence interval		t value	p value		95% confidence interval		t value	p value			
Lumbar erector spinae	-6.299 to 10.966	absence	-0.011	p≥0.05	absence	-7.217 to 11.717	absence	0.716	p≥0.05	absence		
Rectus femoris	-13.241 to 0.407	absence	0.868	p≥0.05	absence	-2.172 to 20.506	absence	0.994	p≥0.05	absence		
(Equipment B)												
Lumbar erector spinae	-0.280 to 1.402	absence	2.071	p≥0.05	absence	0.181 to 1.271	presence	1.253	p≥0.05	absence		
Rectus femoris	-0.391 to 0.919	absence	1.161	p≥0.05	absence	-0.209 to 1.067	absence	-0.164	p≥0.05	absence		
(Equipment C)												
Lumbar erector spinae	-0.711 to 10.231	absence	-0.275	p≥0.05	absence	-4.835 to 5.628	absence	-0.215	p≥0.05	absence		
Rectus femoris	-2.546 to 6.116	absence	-0.347	p≥0.05	absence	-0.094 to 5.648	absence	-0.062	p≥0.05	absence		
Beginner II												
	Fixed bias			Proportional bias			Fixed bias			Proportional bias		
(Equipment A)	95% confidence interval		t value	p value		95% confidence interval		t value	p value			
Lumbar erector spinae	-25.605 to -4.062	presence	-1.738	p≥0.05	absence	-8.944 to 17.944	absence	0.402	p≥0.05	absence		
Rectus femoris	-6.793 to 0.793	absence	-0.983	p≥0.05	absence	-9.542 to 11.376	absence	-1.738	p≥0.05	absence		
(Equipment B)												
Lumbar erector spinae	-0.547 to 0.547	absence	-1.034	p≥0.05	absence	-0.319 to 1.045	absence	1.003	p≥0.05	absence		
Rectus femoris	-0.446 to 0.776	absence	-1.453	p≥0.05	absence	-0.339 to 0.603	absence	-0.042	p≥0.05	absence		
(Equipment C)												
Lumbar erector spinae	-5.173 to 2.397	absence	0.996	p≥0.05	absence	-4.256 to 5.843	absence	0.366	p≥0.05	absence		
Rectus femoris	0.945 to 8.178	presence	0.862	p≥0.05	absence	-1.882 to 9.022	absence	-0.645	p≥0.05	absence		

Table 3. Absolute inter-rater reliability of measurements taken by Expert group and Beginner group (presence or absence of fixed bias and proportional bias), as determined by the Bland–Altman analysis

	Relaxed (R) limb position				Stretched (S) limb position			
	Fixed bias		Proportional bias		Fixed bias		Proportional bias	
Expert group	95% confidence interval		t value	p value	95% confidence interval		t value	p value
(Equipment A)								
Lumbar erector spinae	-11.635 to 10.302	absence	-0.279	$p \geq 0.05$	absence	-27.624 to 6.958	absence	-2.007 $p \geq 0.05$ absence
Rectus femoris	-6.879 to 5.296	absence	1.639	$p \geq 0.05$	absence	-18.401 to 17.151	absence	0.455 $p \geq 0.05$ absence
(Equipment B)								
Lumbar erector spinae	-1.411 to 0.157	absence	-0.977	$p \geq 0.05$	absence	-1.394 to -0.288	presence	-0.919 $p \geq 0.05$ absence
Rectus femoris	-1.659 to -0.584	presence	0.323	$p \geq 0.05$	absence	-1.177 to -0.208	presence	-0.703 $p \geq 0.05$ absence
(Equipment C)								
Lumbar erector spinae	-10.372 to 0.654	absence	-0.625	$p \geq 0.05$	absence	-8.343 to 4.178	absence	-1.65 $p \geq 0.05$ absence
Rectus femoris	-11.883 to -1.008	presence	0.117	$p \geq 0.05$	absence	-7.130 to -0.802	presence	0.816 $p \geq 0.05$ absence
Beginner group								
(Equipment A)	95% confidence interval		t value	p value	95% confidence interval		t value	p value
Lumbar erector spinae	-18.480 to 13.480	absence	-1.878	$p \geq 0.05$	absence	-15.530 to 9.443	absence	-1.15 $p \geq 0.05$ absence
Rectus femoris	-10.331 to 1.081	absence	-0.499	$p \geq 0.06$	absence	-9.595 to 41.845	absence	0.622 $p \geq 0.06$ absence
(Equipment B)								
Lumbar erector spinae	-0.242 to 1.331	absence	1.728	$p \geq 0.05$	absence	-0.314 to 1.535	absence	-0.537 $p \geq 0.05$ absence
Rectus femoris	-0.280 to 1.566	absence	1.398	$p \geq 0.05$	absence	-0.298 to 0.727	absence	2.57 $p < 0.05$ presence
(Equipment C)								
Lumbar erector spinae	-5.596 to 3.018	absence	-0.472	$p \geq 0.05$	absence	-0.584 to 11.690	absence	1.221 $p \geq 0.05$ absence
Rectus femoris	-3.256 to 3.653	absence	-0.732	$p \geq 0.05$	absence	-6.642 to 4.262	absence	-0.248 $p \geq 0.05$ absence

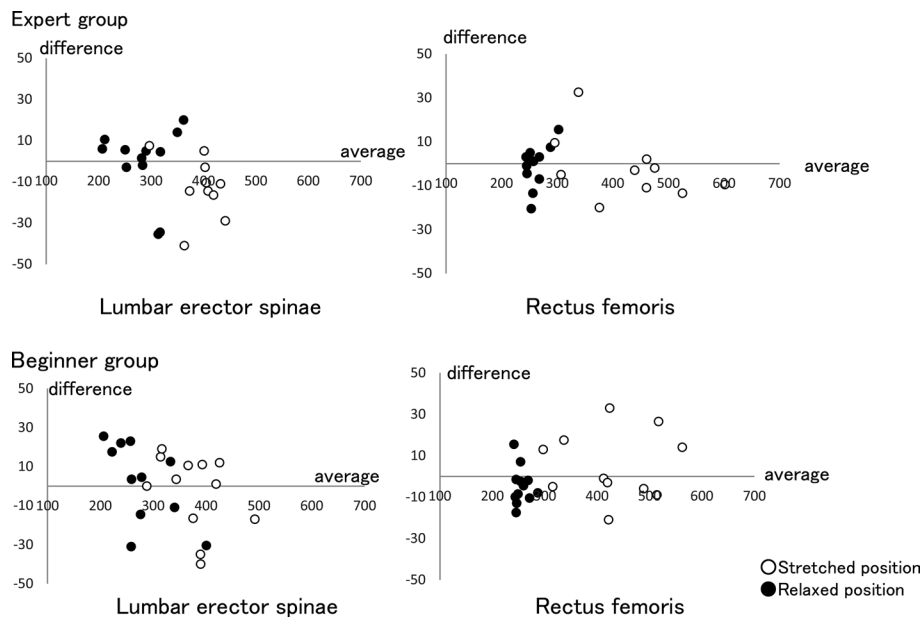


Fig. 2. Bland–Altman plot of measurements taken using Equipment A (MYOTONE-Pro) by Expert group and Beginner group (N/m). The inter-rater reliability of measurements in each test muscle in the relaxed position and stretched position.

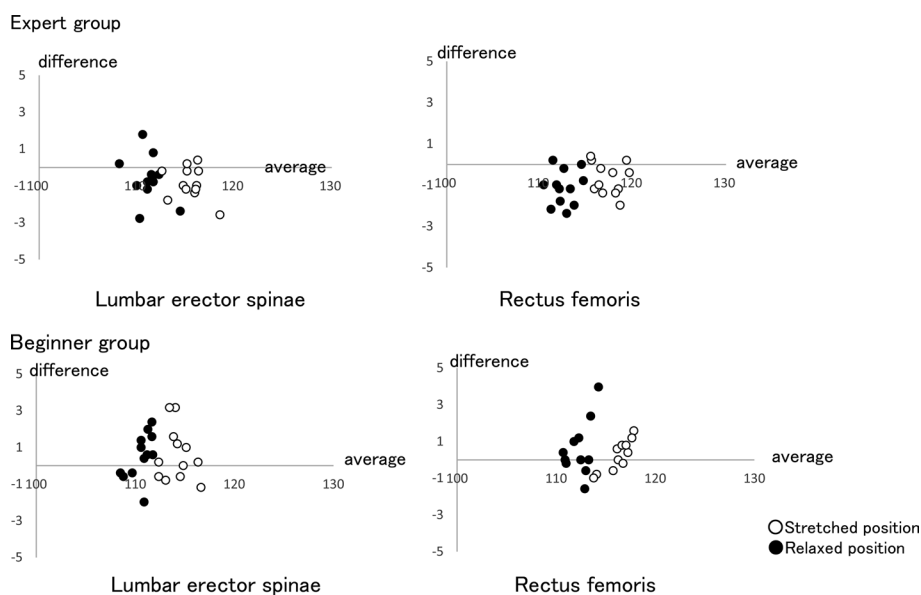


Fig. 3. Bland–Altman plot of measurements taken using Equipment B (PEK-MP) by Expert group and Beginner group (N/m). The inter-rater reliability of measurements in each test muscle in the relaxed position and stretched position.

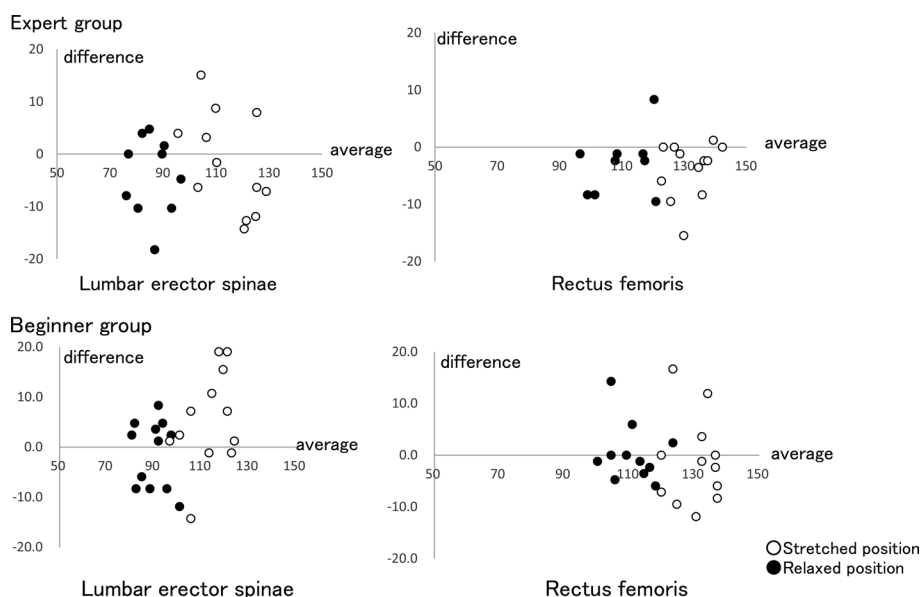


Fig. 4. Bland–Altman plot of measurements taken using Equipment C (NEUTONE TDM-Z2) by Expert group and Beginner group (N/m). The inter-rater reliability of measurements in each test muscle in the relaxed position and stretched position.

may have produced less fixed bias. However, according to previous reports, it is possible that the measurements taken by beginners that lacked proficiency produced measurement errors in the form of random errors and affected measurement reliability. The number of fixed biases in inter-rater reliability in the expert group was higher in the R than in the S position. This was inconsistent with the result of the fixed bias in intra-rater reliability in this study. As the intra-rater reliability of measurements in the expert group improved in the S position compared with the R position, we believe that this improvement was inconsistent between raters. Especially, the inter-rater comparison of measurements taken by the beginner group proved to have favorable trueness but low precision. In contrast, the inter-rater comparison of measurements taken by the expert group was thought to have favorable precision but low trueness. In this study, the participants had 5 min to acclimatize in a resting supine position. One rater took measurements from all tested muscles of one participant in all limb positions using one

Table 4. Differences between muscle hardness measurements and MDC₉₅ values in the relaxed (R) position and stretched (S) positions

	Expert group		Beginner group	
	Differential value	MDC ₉₅	Differential value	MDC ₉₅
(Equipment A)	Hardness (N·m)	Hardness (N·m)	Hardness (N·m)	Hardness (N·m)
Lumbar erector spinae	106.208 ± 37.072	33.836	93.562 ± 9.365	49.295
Rectus femoris	178.958 ± 90.869	18.779	177.583 ± 83.939	17.601
(Equipment B)				
Lumbar erector spinae	3.616 ± 1.195	2.419	3.613 ± 1.195	2.427
Rectus femoris	3.959 ± 1.511	1.657	3.959 ± 1.511	2.847
(Equipment C)				
Lumbar erector spinae	27.072 ± 7.231	17.008	22.759 ± 8.254	13.286
Rectus femoris	22.362 ± 9.461	16.774	16.858 ± 9.366	10.658

tissue hardness meter; the participants had 5 min to acclimatize thereafter before repeating the measurements by measuring another participant and changing the measurement equipment. In these measurements, a randomization chart was used to randomize the order of measurement regarding the combination of tested muscles and measurement equipment. However, no acclimation interval was provided at any time between the measurements of all tested muscles in all limb positions by each rater using one tissue hardness meter until measurements were completed. This might have influenced the results of the presence or absence of fixed bias in inter-rater reliability in the beginner group and the increase in the number of fixed bias in the measurements in the S position taken by the expert group. As beginners are not accustomed to handling equipment during measurement, changing the measurement interval and limb position cost time, making measurements less precise but with good trueness. Experts were accustomed to the measurement equipment, measurement procedure, and changes in the position of participants, and could perform highly proficient and precise measurements. As the experts were highly proficient in performing the measurements, the measurement intervals and time taken to change limb position were short. The shortness of these intervals may have reduced trueness and increased the number of fixed biases. Particularly, the measurement of the rectus femoris muscle includes a self-repositioning movement, in which the participant himself shifts to the Thomas limb position after raters take measurements in the R position (i.e., the supine position); the accuracy of the instructions for this repositioning and how quickly the participant could reposition himself may also have affected the outcomes. Therefore, we understood that, even if a measurement proves to have high relative reliability, studying absolute reliability may help identify problems. It is possible that the opposite phenomenon may have also occurred. Thus, it is necessary to examine relative and absolute reliabilities to investigate measurement reliability. Moreover, we believe that it is necessary to provide a resting acclimation interval after the measurement of each muscle and perform targeted measurements to improve absolute inter-rater reliability. Regarding acclimation time and limb position during measurement, it may be necessary to consider a resting position, in which 3–5 min can easily pass, with the actual limb position used for measurement maintained without the need for repositioning the body. In addition, the measurement surface should be an anti-gravity surface.

The differences determined by subtracting the R from S position measurements exceeded the MDC₉₅ values of all measurement equipment for the expert and beginner groups. In all figures, S position measurements on the x-axis plot are plotted on the right side of R position measurements. Regarding the clinical interpretation of MDC₉₅, Shimoï et al.^{24, 25} indicated that, if the difference obtained through the aforementioned subtraction was less than the MDC₉₅ value obtained from two measurements from the same participant, this discrepancy could be attributed to a measurement error. In contrast, if the difference exceeded the MDC₉₅ value, this discrepancy could be regarded as a “true change” that occurred in the participant, including intervention effects and changes because of aging. We determined that the S position measurements in this study were true changes greater than measurement errors. In the S position, it may be possible to improve the sensitivity of muscle hardness measurements by appropriately stretching each muscle and increasing resting tension before measuring muscle hardness. These findings suggested that it may be possible to produce highly reliable measurements if measurements are taken in the S position.

In conclusion, we studied the absolute intra-rater and inter-rater reliabilities of measurements taken by the expert and beginner groups using the Bland–Altman analysis. Tissue and muscle hardness measurements are important for comparing physical therapy outcomes and an evaluation index of the high necessity for constructing EBM for future physical therapy. The results of our study indicated that there is a need to reconsider the limb position during measurement and the acclimation time provided when taking measurements using a tissue hardness meter. Our experiments also clarified the need to consider both relative and absolute reliabilities. In the future, we will conduct studies to address the aforementioned issues, make progress in the standardization of measurement methods and evaluation methods, and develop educational tools for these methodologies.

Funding

None.

Conflicts of interest

None.

REFERENCES

- 1) Fischer AA: Muscle tone in normal persons measured by tissue compliance. *J Neuro Orthop Med Surg*, 1987, 8: 227–234.
- 2) Fischer AA: Tissue compliance meter for objective, quantitative documentation of soft tissue consistency and pathology. *Arch Phys Med Rehabil*, 1987, 68: 122–125. [[Medline](#)]
- 3) Sanders GE, Lawson DA: Stability of paraspinal tissue compliance in normal subjects. *J Manipulative Physiol Ther*, 1992, 15: 361–364. [[Medline](#)]
- 4) Kawchuk G, Herzog W: The reliability and accuracy of a standard method of tissue compliance assessment. *J Manipulative Physiol Ther*, 1995, 18: 298–301. [[Medline](#)]
- 5) Kawchuk G, Herzog W: A new technique of tissue stiffness (compliance) assessment: its reliability, accuracy and comparison with an existing method. *J Manipulative Physiol Ther*, 1996, 19: 13–18. [[Medline](#)]
- 6) Komiya H, Maeda J, Takemiya T: A new functional measurement of muscle stiffness in humans. *Exerc Sports Physiol*, 1996, 2: 31–38.
- 7) Roberts KL: Reliability and validity of an instrument to measure tissue hardness in breasts. *Aust J Adv Nurs*, 1998, 16: 19–23. [[Medline](#)]
- 8) Arokoski JP, Surakka J, Ojala T, et al.: Feasibility of the use of a novel soft tissue stiffness meter. *Physiol Meas*, 2005, 26: 215–228. [[Medline](#)] [[CrossRef](#)]
- 9) Yano T, Arima Y, Imoto T: Development of simultaneous pain threshold—biological tissue hardness meter “Digital Palpometer” first report. *J Jpn Phys Ther Assoc*, 1998, 9: 33–39.
- 10) Morozumi K, Fujiwara T, Karasuno H, et al.: A new tissue hardness meter and algometer; a new meter incorporating the functions of a tissue hardness meter and an algometer. *J Phys Ther*, 2010, 22: 239–245. [[CrossRef](#)]
- 11) Uchiyama T, Osugi K, Murayama M: Evaluation of muscle hardness using push-in reaction force measurement—impact of isometric contractile force dependence and muscle fatigue—. *J Soc Biomech*, 2006, 18: 219–227.
- 12) Takanashi A, Karasuno D, Shiota K, et al.: Reproducibility of two types of soft tissue hardness meter—a study of reliability—. *Rigakuryoho Kagaku*, 2008, 23: 297–300. [[CrossRef](#)]
- 13) Takanashi A, Karasuno D, Kato M, et al.: Examination of reliability when measuring a simulated soft tissue model with a soft tissue hardness tester. *Rigakuryoho Kagaku*, 2009, 24: 31–34. [[CrossRef](#)]
- 14) Takanashi A, Kawata K, Shiota K, et al.: Evaluation of elasticity using soft tissue hardness tester. *Rigakuryoho Kagaku*, 2011, 26: 667–671. [[CrossRef](#)]
- 15) Watanabe T, Morishita K, Karasuno D, et al.: Examination of new objective measurement method of muscle hardness that reflects performance. *J Appl Biometrol*, 2011, 2: 23–26.
- 16) Morozumi K: Method of measurement of soft tissue hardness. *Rigakuryoho Kagaku*, 2013, 30: 363–369.
- 17) Morozumi K, Ichikawa T, Sugimoto A, et al.: Basic study of spastic muscle evaluation using soft tissue hardness tester. *J Appl Biometrol*, 2013, 4: 43–47.
- 18) Yamakawa M: Principles of ultrasound elastography. *J Soc Biomech*, 2016, 40: 73–78. [[CrossRef](#)]
- 19) Uchiyama T: Special article: Hardness of flexible objects. *J Soc Biomech*, 2016, 40: 72. [[CrossRef](#)]
- 20) Murayama M: Significance of muscle hardness evaluation using push-in reaction force measurement. *J Soc Biomech*, 2016, 40: 79–84. [[CrossRef](#)]
- 21) Morozumi K, Hanaoka M, Ichikawa T, et al.: Validity of quantitative evaluation method of spastic muscle using portable tissue hardness meter. *Josai International University bulletin*, 2017, 25: 41–57.
- 22) Morozumi K, Hanaoka M, Ichikawa T, et al.: Relationship between spastic muscle hardness measurement using a portable tissue hardness meter and spasticity evaluation scale—a study using multiple regression analysis in patients with spastic paralysis. *Josai International University bulletin*, 2018, 26: 51–61.
- 23) Morozumi K, Morishita K, Aoki M, et al.: Effects of proficiency in muscle hardness measurement, different instruments, and positioning on the intra-rater and the inter-rater reliability. *Rigakuryoho Kagaku*, 2021, 36: 689–698. [[CrossRef](#)]
- 24) Shimoi T: Absolute reliability of assessment. *Rigakuryoho Kagaku*, 2011, 26: 451–461. [[CrossRef](#)]
- 25) Shimoi T, Tani H: A study of the absolute reliability of two types of tandem gait tests using minimum detectable change. *Rigakuryoho Kagaku*, 2010, 25: 49–53. [[CrossRef](#)]
- 26) Shimoi T, Tani H: A study of intra-rater and inter-rater reliability of tandem gait test using Bland-Altman analysis. *Rigakuryoho Kagaku*, 2008, 23: 625–631. [[CrossRef](#)]
- 27) Arima Y: Objectification and induration of biological surface hardness. *J Soc Biomech*, 2016, 40: 85–90. [[CrossRef](#)]