CHAPTER 2

LITERATURE REVIEW

2.1 Intrarater and interrater reliability of lateral abdominal muscles thickness measurement using B-mode ultrasound imaging

2.1.1 Introduction

Ultrasound imaging is currently used extensively for assessment and facilitation of deep muscles in rehabilitation program [24]. RUSI is consider to be a safe, cost-effective, and accessible method for visualizing and measuring the deep muscles of the trunk [14, 53, 63, 64, 66]. The value of ultrasound imaging from a rehabilitative perspective is that it allows for dynamic study (real-time images) of muscle groups as they contract. In addition to its clinical utility, numerous studies [7, 14, 19, 54, 61, 64, 66] have shown that ultrasound imaging is both a valid and reliable method to ascertain direct visualization of muscle size (through static quantitative measurements of muscle width, length, depth, cross-sectional area, or volume) and hence can be used as an indicator of muscle activity [7, 14, 19, 54, 61, 64, 66].

The reliability of RUSI as an instrument for measuring muscle thickness has been reported in some studies [13, 49, 61, 79]. Although the validity of measuring changes in muscle thickness as an indicator of TrA and IO muscle activation using RUSI has been addressed [14, 26], assessment of techniques to improve reliability and to decrease measurement error remains unknown. Previous assessment of intrarater reliability [7, 49, 61, 66, 79] has found acceptable levels of reliability.

However, the standard error of the measurement (SEM) remains relatively high compared to the resting thickness of the lateral abdominal muscles. For example, the reported 0.31mm SEM for the TrA would require a 29.5% change (based on a 2.1mm TrA muscle thickness) in muscle thickness to be 95% confident a change has occurred. These high SEM values would limit the ability to detect a change in muscle thickness with RUSI in longitudinal studies. Springer et al [80] demonstrated that the 50% decrease in SEM obtained when an average of three measures were assessed for TrA. However, the influence of performance variability of the LAM measurement on response stability values (ie, SEM) remains unknown. Further analysis of interrater reliability, specifically techniques to decrease the SEM, is required. Thus, the purposes of this study were to determine intraimage intrarater reliability, within and between day intrarater reliability and interrater reliability using B-mode ultrasound.

2.1.2 Methods

2.1.2.1 Subjects

Ten healthy female subjects (aged 22.40 $\pm$ 0.52 years) participated in the study. Subjects with history of low back pain in past three months, significant spinal deformity, history of lumbar surgery, history of neuromuscular or joint disease or neurological conditions affecting the trunk or pregnancy were excluded. The study was approved by Human Experimental Committee of Faculty of Associated Medical Sciences, Chiang Mai University. Informed consent was obtained from all subjects.

2.1.2.2 Procedure

Ultrasound Measurements Ultrasound images of the right lateral abdominal wall were obtained using an ultrasound scanner (Toshiba, Famio 8, SSA-530A) in B-mode with a 12-MHz linear array transducer. Subjects were positioned on a plinth in crook lying with a pillow under their head and the knees (Figure A1-1). Ultrasonic gel was interposed between the transducer and the skin. The transducer was placed in a transverse plane just superior to the right iliac crest along the axillary line [(7)]. Image of LAM and the thickness measurements (Figure A1-2) were obtained following techniques as described in Appendix 1. Interrater reliability was assessed by two investigators (P.S. and U.P.). Intrarater reliability was determined for intraimage and interimage reliability. Interimage, intrarater reliability was measured from two images. First image was performed on day one and second image was repeated one week later. Intraimage, intrarater reliability was determined from three measurements of the first image. Images were stored and measured off line using National Institutes of Health (NIH) ImageJ software (Version 1.38t) (http://rsb.info.nih.gov/nih-image/). All images were coded and randomly selected for measurement. To minimize recall bias, the examiner did not review the measurements value after the completion of the first measurement session.

2.1.2.3 Statistical analysis

The analysis of reliability across three repeated measurements of the same image (intraimage) was calculated by intraclass correlation coefficient (ICC) model 3 ($ICC_{3,1}$). The analysis of reliability across the three repeated images (interimage) was calculated from the average of three measurements from each image using $ICC_{3,3}$. Interater reliability was calculated using $ICC_{2,3}$. Response stability was

calculated using standard error of the measurement (SEM = pooled SD x $\sqrt{1-ICC}$) and the coefficient of variation method error (CV_{ME}). The CV of method error was calculated for test-retest reliability using the formula $CV_{ME} = \frac{2ME}{X1+X2}x100$. In this formula, method error (ME) is a measure of the discrepancy between test and retest scores and is calculated as follows after determining the mean and standard deviation of the difference between the repeat tests: $ME = \frac{SDdiff}{\sqrt{2}}$. Larger difference scores reflect greater measurement error. X1 and X2 represent the measurements from tests at the first visit and the following visit, respectively. The values of the CV of method error reflects the amount of variation in the difference scores in the test and retest measures and is relative to the size of the mean differences. The minimal detectable change (MDC_{95} = SEM x $\sqrt{2}$ x 1.96) were calculated for each ultrasound measurement taken. Bland and Altman tests [81] were used to calculate mean difference (bias) between measures with 95% confidence interval (95% CI) and 95% limit of agreements (95% LOA = $\pm$ 2SD) were calculated.

2.1.3 Results

The descriptive data for subjects enrolled in the study are provided in Table 2-1. Mean and SD of LAM thickness is shown in Table 2-2. Reliability coefficients with corresponding 95%CI, SEM, MDC_{95}, bias and LOA are presented in Table 2-3 for intraimage, intrarater reliability, Table 2-4 for interimage, intrarater within day reliability, Table 2-5 for interimage, intrarater between day reliability and Table 2-6 for interrater reliability. The $ICC_{3,1}$ intrarater reliability results were greater than 0.99 for intraimmage (Table 2-3). The interimage reliability ranged from 0.95 to

0.99 for withinday and from 0.86 to 0.96 for between day (Table 2-4 and Table 2-5). The interrater reliability was greater than 0.88 (Table 2-6). The results indicate good to excellent reliability (ICCs ranged from 0.88 to 0.99).

SEM was calculated to determine the precision. Overall SEM of LAM thickness were less than 0.1 mm for intraimage, intrarater estimates, less than 0.2 mm for interimage within day, less than 0.4 for interimage between day and less than 0.9 mm for interrater estimates.

MDC_{95} represented the minimal change in thickness that a true change occurred. MDC_{95} of LAM thickness was less than 0.2 mm for intraimage estimates, less than 0.7 mm for interimage within day estimates, less than 1.3 mm for interimage between day estimates and less than 2.5 mm between investigators. Bias estimates were small and estimate ranged between 0.1 and 0.2 mm for overall LAM intrarater measurements and 0.1 and 0.7 mm for interrater measurements.

Table 2-1 Descriptive data of subjects (n=10)

	Mean ± SD	**Range**
Age (yr)	22.40 ± 0.52	22 - 23
Weight (Kg)	50.79 ± 5.29	44- 60
Height (cm)	160.16 ± 3.90	153.90 – 167.00
BMI (kg/m2)	19.85 ± 2.44	16.77 – 24.03

Table 2-2 Mean and SD of LAM thickness

LAM thickness (mm)	Mean ± SD	Range
TrA	2.13 ± 0.60	1.21 – 3.62
IO	6.53 ± 1.26	3.28 – 9.14
EO	4.45 ± 0.83	2.41 – 6.72
IEO	11.71 ± 1.99	6.72 – 16.72
Total	14.58 ± 2.69	8.45 – 21.90

Table 2-3 The intraimage, intrarater reliability

Muscles	$ICC_{3,1}$(95%CI)	SEM (mm)	$\%CV_{ME}$	MDC_{95} (mm)	Bias (95%CI) ± 95%LOA (mm)
TrA	0.995(0.981-0.998)	0.04	3.01	0.12	0.07 (0.00-0.13) ± 0.18
IO	0.998(0.992-0.998)	0.05	1.09	0.14	-0.02 (-0.01-0.05) ± 0.20
EO	0.994(0.976-0.998)	0.06	2.11	0.17	-0.02 (-0.11-0.07) ± 0.26
IEO	0.998(0.992-0.999)	0.08	1.02	0.22	-0.001 (-0.12-0.12) ± 0.32
Total	0.999(0.997-0.999)	0.06	0.57	0.16	-0.04 (-0.12-0.05) ± 0.22

Table 2-4 The interimage, intrarater reliability within day

Muscles	$ICC_{3,3}$(95%CI)	SEM (mm)	$\%CV_{ME}$	MDC_{95} (mm)	Bias (95%CI) ± 95%LOA (mm)
TrA	0.988(0.951-0.9970	0.07	4.54	0.19	-0.03 (-0.13-0.07) ± 0.26
IO	0.977(0.907-0.994)	0.18	3.89	0.50	-0.10 (-0.36-0.15) ± 0.72
EO	0.955(0.817-0.988)	0.17	5.56	0.47	-0.07 (-0.31-0.17) ± 0.68
IEO	0.980(0.918-0.995)	0.26	3.27	0.72	-0.14 (-0.52-0.24) ± 1.06
Total	0.990(0.960-0.997)	0.23	2.36	0.64	-0.16 (-0.49-0.18) ± 0.94

Table 2-5 The interimage, intrarater reliability between day

Muscles	$ICC_{3,3}$(95%CI)	SEM (mm)	$\%CV_{ME}$	MDC_{95} (mm)	Bias (95%CI) ± 95%LOA (mm)
TrA	0.969(0.874-0.992)	0.12	7.92	0.33	-0.08 (-0.25-0.09) ± 0.48
IO	0.919(0.675-0.980)	0.31	6.77	0.86	0.18 (-0.25-0.62) ± 1.22
EO	0.869(0.471-0.967)	0.29	9.43	0.80	0.06 (-0.34-0.46) ± 1.12
IEO	0.969(0.877-0.992)	0.32	4.00	0.89	0.08 (-0.38-0.53) ± 1.28
Total	0.960(0.839-0.990)	0.48	4.88	1.33	-0.08 (-0.78-0.61) ± 1.94

Table 2-6 The interrater reliability

Muscles	$ICC_{2,3}$(95%CI)	SEM (mm)	$\%CV_{ME}$	MDC_{95} (mm)	Bias (95%CI) ± 95%LOA (mm)
TrA	0.887(0.546-0.972)	0.17	10.76	0.47	-0.17 (-0.40-0.06) ± 0.66
IO	0.942(0.768-0.985)	0.34	7.12	0.94	0.17 (-0.31-0.65) ± 1.34
EO	0.891(0.561-0.972)	0.29	8.67	0.80	-0.22 (-0.63-0.18) ± 1.14
IEO	0.955(0.818-0.988)	0.45	5.30	1.25	-0.09 (-0.74-0.56) ± 1.82
Total	0.909(0.632-0.977)	0.90	8.26	2.49	-0.78 (-2.00-0.48) ± 3.60

2.1.4 Discussion

This study evaluated the intrarater and interrater reliability in obtaining RUSI thickness measurements of LAM at rest. The results of this study indicate that B-mode ultrasound used in this study is reliable in detecting the thickness of LAM in normal subjects.

Previous studies [61, 66] reported moderate to good reliability of M-mode ultrasound of abdominal muscles in supine, standing and walking position in asymptomatic subjects (ICCs 0.88 to 0.94). ICCs for changes in thickness between supine and standing were 0.78, and between supine and walking were 0.48. Chitchley and Coutts [13] showed reliable measurement for B-mode ultrasound in four-point kneeling at rest and during an abdominal hollowing maneuver (ICCs 0.94 to 0.98). Norasteh et al [79] reported the reliability of B-mode ultrasound of abdominal muscles in symptomatic subjects (ICCs 0.72 to 0.97).

Springer et al [80] demonstrated that the use of an average of three measurements was able to decrease the SEM by over 50% in interrater reliability of

TrA muscle thickness (SEM 0.13-0.35 mm). It should be noted that measurement in the current study was different from those used in the literature. The current study explained measurement in all of LAM. The study used average of three measurements from each image which was able to decrease SEM. The intraimage, intrarater reliability of TrA (SEM, 0.04 mm and 3 %CV_{ME}) and the interimage, intrarater reliability of TrA (SEM, 0.12 mm and 7.92 %CV_{ME}) measured in the current study were lower than Teyhen et al [7] who measured intraimage, intrarater SEM 0.13 mm and 5% CV_{ME} and interimage, intrarater SEM 0.31 mm and 11% CV_{ME}. The interrater reliability of TrA (SEM 0.17 mm and 10.76 %CV_{ME}) measured in current study was close to Springer et al [80] which demonstrated SEM of TrA muscle of 0.13 mm.

The present study showed high intrarater reliability across day for LAM (ICCs 0.87 to 0.97) and SEM was 0.12 to 0.32 mm. It is helpful to determine the error related to measurements on different day for monitoring recovery or training effect on muscle thickness. Therefore, any changes attributed to signs of recovery or training should be outside the margin of measurement error.

Regardless of where the specific bias occurred, the actual difference between measurements were very small both intrarater and interrater. The interpretation of reliability coefficients should consider both precision of the measured variable and how the measure be ultimately used. RUSI measurement can be used clinically to make decision on muscle thickness change regarding to intervention such as spinal stabilization exercise.

A limitation of this study is that the measurements were only conducted at rest. Further studies could examine the reliability of performing the thickness measurements on voluntary contraction and in different populations. In

addition, both investigators were novice ultrasound image experience. Some evidence suggests that the reliability of RUSI measurement may differ depending on user experience [82].

In conclusion, the results of this study indicate that measurement of LAM thickness with B-mode ultrasound can be performed reliably. This might help to assess LAM muscle thickness in clinical setting. The establishment of reliable measures of the LAM thickness is the first step towards investigating muscle size and function.

2.2 Intrarater and interrater reliability of assessment of lumbar multifidus muscle cross-sectional area assessment using ultrasound imaging

2.2.1 Introduction

Lumbar multifidus muscle plays an important role in segmental stabilization of lumbar spine and prevention of recurrence after episode of LBP [50, 83, 84]. The CSA of a muscle is directly related to its ability to produce force. Thus, CSA could be used as an indirect assessment of muscle strength. RUSI provides a safe and suitable tool in clinical setting to assess the CSA of LM and other deep muscles that are the target of stabilization exercises. Validity of LM muscle CSA measurements has been established by comparing measurements with magnetic resonance imaging (MRI) [85].

The error associated with repeated measurements over time is a critical issue for any measurement. Hides et al [69] examined the CSA and linear measurements using RUSI for the repeatability in normal young adults. The results showed that measurements were repeatable between day (CV=6%) and between scans

(CV=4.9%). Therefore, they were more appropriate for clinical used. Hides et al [85] report a 3.58% CV for LM muscle CSA measurements when made by the same skilled RUSI on 2 different day.

Pressler et al [86] reported a reasonable between-day intrarater reliability for S1 multifidus muscle in a physical therapist, newly trained in RUSI (ICC = 0.80 for the right and 0.72 for the left). The between-day SEM was 0.37 cm^2.

Keisel et al [16] conducted on-screen measurements of LM and reported good intrarater reliability (ICC = 0.85). Images of LM were also saved for later measurement by two raters, and high interrater reliability was founded (ICC = 0.95).

Wallwork et al [87]conducted an intrarater and interrater reliability of LM thickness assessment at L2-3 and L4-5 zygaphophyseal joints using RUSI. It was founded that a novice and an experienced investigator produced high intrarater reliability at L2-3 zygaphophyseal joints (ICC= 0.89 and 0.94) and at L4-5 zygaphophyseal joints (ICC= 0.88 and 0.95). An average of three trials produced higher interrater reliability for both vertebral levels (ICC > 0.96).

Lumbar multifidus has been assessed in healthy subjects and low back pain patients using RUSI and conducted by an experienced investigator. Several studies have reported the reliability of LM measurements using RUSI [16, 54, 69, 85, 86]. Most of these studies have shown very high reliability (ICC > 0.9) and good precision (SEM < 3.7 mm). However, few studies have examined the reliability of novice investigator. The purpose of this study was to examine intrarater and interrater reliability of an experienced and a novice investigator performing assessment CSA of LM using RUSI in healthy subjects at rest.

2.2.2 Methods

2.2.2.1 Subjects

Eight volunteers (5 female, 3 male) without a history of low back pain were participated in this study. The mean (± SD) age, height, weight and body mass index (BMI) of the subjects were 38.75 ± 10.17 years, 172.25 ± 5.95 cm, 67.94 ± 12.05 kg, and 22.89 ± 3.77 kg/m^2, respectively. Informed consent was obtained from all subjects.

2.2.2.2 Procedure

Ultrasound imaging of LM muscle CSA was measured using 5-MHz curvilinear transducer (Phillip Ultrasound). LM muscle CSA was assessed bilaterally at L2, L3, L4 and L5 vertebral levels in prone position (Figure A1-3). The ultrasound measurements were performed by one experienced and one novice investigators. The experienced operator had 20 years experience in assessing muscles using RUSI. Training of the novice rater was provided by the experienced investigator. In addition, the novice rater and the experienced operator measured 5 subjects prior to commencing the study. Both raters were experienced physical therapists familiar with clinical assessment of the lumbar spine and location of bony landmarks in the region. Each rater manually palpated the spinous processes of L2-L5 lumbar vertebral levels and marked on the skin with a pen. Raters were allowed to confirm their own skin markings using RUSI in longitudinally along the midline of the lumbar spine. The ultrasound images were taken from L2-L5 with subjects in a relaxed state and images were stored for offline analysis (Figure A1-4). Images of LM and CSA measurements were obtained following technique as described in Appendix 1. All skin markings were removed with alcohol after the first rater conducted

measurements, and the subject stood up and moved around. This ensured that each rater independently established anatomical landmarks and subject positioning.

The program Image J was used to calculate the CSA of the LM muscle (Version 1.38t) (http://rsb.info.nih.gov/nih-image/) at the L2-L5 vertebral levels.

2.2.2.3 Statistical analysis

The measurement was carried out at each level by experienced and three times by novice investigator. The first trial was used to determine the interrater reliability. Intrarater (model 3) and interrater (model 2) reliability were assessed by calculation of an F statistic (Analysis of variance), the ICC, and the SEM. The minimal detectable change (MDC_{95} = SEM x $\sqrt{2}$ x 1.96) were calculated for each ultrasound measurement taken. Bland and Altman tests were used to calculate mean difference (bias) between measures with 95% confidence interval (95% CI) and 95% limit of agreements (95% LOA = $\pm$ 2SD) were calculated.

2.2.3 Results

The mean value of LM muscles CSA were shown in Table 2-7. The intrarater reliability was demonstrated across repeated measurement of the same image (Table 2-8) and across three images in novice (Table 2-9). The Analysis of variance indicates no systemic difference in CSA of LM values across measurements for both novice and experienced investigators (all F = 0.06-5.20, $p > 0.05$). Interrater reliability was founded low ICC for single measurement at L2, L3, L4 and L5 vertebral levels ($ICC_{2,1}$ ranged 0.68 to 0.76, 95%CI ranged 0.03 to 0.95). Thus, Interrater reliability based on the average of 3 measures per rater was calculated (Table 2-10).

SEM of LM muscles CSA measurements were less than 0.2 cm^2 for same image both an experienced and a novice investigator. SEM of LM CSA measurements between investigator were less than 0.6 cm^2. Thus, overall SEM of LM measurements in this study showed good precision. MDC_{95} were less than 0.5 cm^2 for CSA measurements across images and less than 1.8 cm^2 for between investigators. Bias estimates were small and close to zero and did not result in poor reliability when using average of three measures.

Table 2-7 Average CSA of LM (mean $\pm$ SD)

Vertebral level	**Experienced investigator**		**Novice investigator**	
	Left	**Right**	**Left**	**Right**
L2 (cm^2)	1.87 $\pm$ 0.63	1.88 $\pm$ 0.66	2.13 $\pm$ 0.45	2.21 $\pm$ 0.51
L3 (cm^2)	3.07 $\pm$ 1.07	3.10 $\pm$ 1.10	3.39 $\pm$ 0.90	3.54 $\pm$ 0.88
L4 (cm^2)	5.70 $\pm$ 1.49	5.77 $\pm$ 1.61	4.92 $\pm$ 1.01	5.06 $\pm$ 1.07
L5 (cm^2)	6.80 $\pm$ 1.27	6.76 $\pm$ 1.28	7.18 $\pm$ 0.92	7.24 $\pm$ 0.91

Table 2-8 Intrarater reliability across repeated measurement of the same image (intraimage reliability)

Vertebral level	Left				Right			
	ICC 3,1	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2)	ICC 3,1	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2)
experienced								
L2	0.99 (0.99-0.99)	0.03	0.08	-0.01 (-0.09-0.07) ± 0.20	0.99(0.99-0.99)	0.03	0.08	0.01 (-0.10-0.11) ± 0.26
L3	0.94(0.83-0.99)	0.25	0.69	-0.06 (-0.22-0.10) ± 0.38	0.99 (0.98-0.99)	0.06	0.17	-0.04 (-0.02-0.12) ± 0.38
L4	1.00 (0.99-1.00)	0.03	0.08	-0.01 (-0.31-0.30) ± 0.72	0.99 (0.99-1.00)	0.03	0.08	-0.02 (-0.25-0.20) ± 0.54
L5	0.99 (0.99-1.00)	0.03	0.08	-0.03 (-0.09-0.04) ± 0.16	1.00 (0.99-1.00)	0	0.00	-0.02 (-0.13-0.09) ± 0.26
Novice								
L2	0.97(0.90-0.99)	0.07	0.19	0.02 (-0.03-0.07) ± 0.12	0.96 (0.89-0.99)	0.08	0.22	0.01 (-0.02-0.05) ± 0.08
L3	0.97 (0.91-0.99)	0.15	0.42	-0.12 (-0.49-0.24) ± 0.88	0.98(0.93-0.99	0.12	0.33	-0.04 (-0.13-0.04) ± 0.20
L4	0.95 (0.85-0.99)	0.23	0.64	-0.01 (-0.03-0.02) ± 0.06	0.97 (0.90-0.99)	0.18	0.50	-0.01 (-0.04-0.03) ± 0.08
L5	0.99 (0.98-0.99)	0.06	0.17	-0.06 (-0.10- -0.01) ± 0.10	0.99 (0.97-0.99)	0.08	0.22	-0.03 (-0.05-0.00) ± 0.06

Table 2-9 Intrarater reliability across 3 images in novice (interimage reliability)

Vertebral level	Left				Right			
	ICC 3,1	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2)	ICC 3,1	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2))
L2	0.90 (0.70-0.98)	0.14	0.39	0.02 (-0.14-0.18) $\pm$ 0.40	0.90 (0.68-0.97)	0.18	0.50	-0.13 (-0.40-0.14) $\pm$ 0.64
L3	0.98 (0.93-0.99)	0.13	0.36	0.07 (-0.04-0.17) $\pm$ 0.24	0.99 (0.96-0.99)	0.09	0.25	-0.05 (-0.28-0.08) $\pm$ 0.30
L4	0.96 (0.88-0.99)	0.19	0.53	0.03 (-0.22-0.28) $\pm$ 0.60	0.97 (0.91-0.99)	0.18	0.50	0.01 (-0.20-0.20) $\pm$ 0.48
L5	0.95 (0.84-0.99)	0.2	0.55	0.06 (-0.08-0.20) $\pm$ 0.34	0.98 (0.94-0.99)	0.12	0.33	-0.04 (-0.18-0.11) $\pm$ 0.34

Table 2-10 Interrater reliability based on the average of three measures per rater

Vertebral level	Left				Right			
	ICC 2,3	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2)	ICC 2,3	SEM (cm^2)	MDC_{95} (cm^2)	Bias (95%CI) + 95%LOA (cm^2)
L2	0.81 (0.16-0.96)	0.29	0.80	0.22 (-0.13-0.57) $\pm$ 0.84	0.85 (0.20-0.97)	0.24	0.67	0.27 (-0.02-0.56) $\pm$ 0.70
L3	0.86 (0.38-0.97)	0.46	1.28	0.33 (-0.22-0.88) $\pm$ 1.32	0.81 (0.15-0.96)	0.52	1.44	0.42 (-0.19-1.04) $\pm$ 1.48
L4	0.78 (-0.07-0.95)	0.66	1.83	-0.76 (-1.55-0.03) $\pm$ 1.90	0.81 (0.07-0.96)	0.66	1.83	-0.72 (-1.50-0.07) $\pm$ 1.88
L5	0.87 (0.34-0.97)	0.48	1.33	0.46 (-0.11-1.03) $\pm$ 1.36	0.85 (0.24-0.97)	0.51	1.41	0.50 (-0.11-1.11) $\pm$ 1.46

2.2.4 Discussion

Results from this study showed that an experienced and a novice investigator were both able to reliably measure lumbar multifidus muscle CSA at L2,L3,L4 and L5 vertebral levels using RUSI. The interpretation of reliability estimates was suggested by Portney and Watkins [88] that coefficients below 50 represent poor reliability, 0.50-0.75 represent moderate reliability and above 0.75 represent good reliability. An experienced and a novice investigator exhibited high intrarater reliability in same image when a single trial per rater was used (overall $ICC_{3,1} > 0.95$). Intrarater reliability across images in a novice was also high ($ICC_{3,1} > 0.9$). Using an average of three measures per rater produced good interater reliability ($ICC_{2,3}$ ranged 0.78 to 0.87; 95%CI -0.07 to 0.97). This was supported by previous studies [80, 87, 89] that an average of three trials produce very high interrater agreement on the measurement of LM thicknesses. Wallwork et al [87] reported high interrater reliability of LM thickness level ($ICC_{2,3} = 0.96$ at the L2-3 and 0.97 at the L4-5). In addition, the novice and experienced operator produced reliable intrarater measurements ($ICC_{3,1} = 0.89$ and 0.94 at L2-3; 0.88 and 0.95 at L4-5).

In addition to the finding of high intrarater and interrater reliability for both an experienced and a novice investigator, the reported SEM values were low (<0.66 cm^2). This is important for clinical application when assessing changes in muscle CSA. A low SEM value could be used to detect a real change that would exceed measurement error. The consistently low SEM values also indicate low measurement error. Wallwork et al [87] reported SEM of interrater estimates LM thickness 0.06 cm at L2-3 and 0.5 cm at L4-5. SEM of a novice assessor was 0.11 cm

at L2-3 and 0.09 cm at L4-5. An experienced showed SEM 0.09 cm at L2-3 and 0.06 cm at L4-5. However, there was no study that reported the SEM of LM CSA.

The present study also has implications for the responsiveness of LM muscles CSA measures during rehabilitation. A value of LM muscles CSA larger than the MDC confirmed a real change occurred. Moreover, the specific bias occurred in this study, the actual differences between investigators were very small and not resulted in poor reliability.

Before the measurement of LM muscles CSA can be adopted in clinical and research situations, it is important for potential investigators to establish that they can perform the measurement reliably. This requires investigators to standardize measurement protocols. In this study, great care was taken with subject positioning and accurate location of anatomical landmarks. Four vertebral levels were selected for measurement as the morphology of the LM muscle is known to vary over vertebral levels, and increase significantly in size at each vertebral level on progression caudally [(85)]. This anatomical feature allowed for a further check of the reliability of the investigators, as accurate relocation of vertebral levels was therefore required. LM measurements of the L2 to L5 vertebral levels require accurate identification of the lamina and border between LM and longissimus muscle, which are more difficult to define. Stokes et al [(54)] suggested that asking the subject to perform a gentle lift of the ipsilateral leg which will clarify the lateral border of LM from longissimus muscle. Further consideration in measurement was to define the precise location of the cursors. In this study, the inside border of the LM muscle was selected. Another important factor that can influence the measurement is transducer

pressure. If the assessor uses too much pressure, the CSA of the muscle will be decreased.

A limitation of this study is that the measurements were only conducted at rest. Future studies could examine the reliability of performing the measurements on voluntary contraction and in different subject populations. Also, the subject number was small and the reliability of measurements across days was not tested. If a measurement is to be repeated to monitor recovery, it is important that the error related to measurements on different days is known, so that any changes attributed to signs of recovery are shown to be outside the margin of measurement error.

In conclusion, CSA of LM muscle using RUSI were measured in transverse section by an experienced and a novice investigator. The result of present study concluded that an experienced and a novice assessor were both able to reliably perform measurement of CSA of LM muscle using RUSI. An average of three trials produced good interrater reliability.