Title: Measuring the morphological characteristics of thoracolumbar fascia in ultrasound images: an inter-rater reliability study.

Authors: Kyra De Coninck, Karen Hambly, John W Dickinson and Louis Passfield.

Author affiliations: School of Sport and Exercise Sciences
University of Kent
Medway Building,
Chatham Maritime
Chatham, Kent
ME4 4AG, UK

Author Email: K.De-Coninck@kent.ac.uk
K.Hambly@kent.ac.uk
J.W.Dickinson@kent.ac.uk
L.Passfield@kent.ac.uk

Corresponding author: Kyra De Coninck
Address: School of Sport and Exercise Sciences
University of Kent
Medway Building,
Chatham Maritime
Chatham, Kent
ME4 4AG, UK

Phone: +44 1634 888809
Email: K.De-Coninck@kent.ac.uk

Word Count: 3710 words main body, 288 words abstract <350 words

ABSTRACT

BACKGROUND: Chronic lower back pain is still regarded as a poorly understood multifactorial condition. Recently, the thoracolumbar fascia complex has been found to be a contributing factor. Ultrasound imaging has shown that people with chronic lower back pain demonstrate both a significant decrease in shear strain, and a 25% increase in thickness of the thoracolumbar fascia. There is sparse data on whether medical practitioners agree on the level of disorganisation in ultrasound images of thoracolumbar fascia. The purpose of this study was to establish inter-rater reliability of the ranking of architectural disorganisation of thoracolumbar fascia on a scale from 'very disorganised' to 'very organised'.

METHODS: An exploratory analysis was performed using a fully crossed design of inter-rater reliability. Thirty observers were recruited, consisting of 21 medical doctors, 7 physiotherapists and 2 radiologists, with an average of 13.03 ± 9.6 years of clinical experience. All 30 observers independently rated the architectural disorganisation of the thoracolumbar fascia in 30 ultrasound scans, on a Likert-type scale with rankings from 1 = very disorganised to 10 = very organised. Internal consistency was assessed using Cronbach's alpha. Krippendorff's alpha was used to calculate the overall inter-rater reliability.

RESULTS: The Krippendorf's alpha was .61, indicating a modest degree of agreement between observers on the different morphologies of thoracolumbar fascia.The Cronbach's alpha (0.98), indicated that there was a high degree of consistency between observers. Experience in ultrasound image analysis did not affect constancy between observers (Cronbach's range between experienced and inexperienced raters: 0.95 and 0.96 respectively).

CONCLUSIONS: Medical practitioners agree on morphological features such as levels of organisation and disorganisation in ultrasound images of thoracolumbar fascia, regardless of experience. Further analysis by an expert panel is required to develop specific classification criteria for thoracolumbar fascia.

Keywords: inter-observer reliability; thoracolumbar fascia; ultrasound imaging

Background

A growing body of evidence supports the notion that the thoracolumbar fascia, an anatomical structure consisting of layers of dense connective tissue in the lumbar area of the trunk, is clinically important in people with chronic lower back pain [1–8]. The thoracolumbar fascia has been shown to play an important role in force transmission between lower limbs and trunk in both ex-vivo cadaver studies [9, 10] and in-vivo research during walking [11, 12]. Subcutaneous fascial bands have been found to mechanically link the skin, subcutaneous layers and deeper muscles. The differences in morphological characteristics of subcutaneous fascial planes may reflect how mechanical forces are distributed across various tissues [13] . However, what is not clear, is whether medical practitioners are able to agree on these different morphological features in ultrasound images of thoracolumbar fascia.

The architecture of the thoracolumbar fascia is complex, it consists of layers of dense collagenous connective tissue, interspersed with loose connective tissue which allows the dense layers to slide and hence play a role in trunk mobility. The thoracolumbar fascia is continuous with the aponeuroses of major trunk muscles which are instrumental in movement and vertebral control [8, 9]. It has been hypothesised that fibrosis, densification and thickening in the thoracolumbar fascia may be the result of an inflammatory response or soft tissue injury [1, 14–17]. For instance, a recent animal study demonstrated that an induced soft tissue injury in the lumbar region, when combined with movement restriction, lead to fibrosis, and significant thickening of thoracolumbar fascia[18]. An earlier pioneering ultrasound based human study concluded that the thoracolumbar fascia in people with chronic lower back pain demonstrated 25% greater thickness compared to a matched control group [4]. A follow-up investigation found that thoracolumbar fascia shear strain during passive trunk flexion, was reduced in people with chronic lower back pain by 56% [19]. In both aforementioned studies, Langevin's research team found significant differences not only in fascial thickness and echogenicity, but also in disorganisation of the architecture of the connective tissues of people with chronic lower back pain. Even though the clinical relevance of fascial tissues has been established [20], to date no classification of thoracolumbar fascia has been developed. In order to develop a classification system, a level of inter-observer reliability of the different types of architecture of thoracolumbar fascia needs to be established.

The aim of this study was to determine the inter-rater reliability for the rating of morphological characteristics of thoracolumbar fascia in ultrasound images, on Likert-type scale, by a range of clinicians.

Methods

Participants

The study was approved by the University of Kent's Ethics Committee and conducted in compliance with the Helsinki Declaration. Informed consent was obtained from all participants.

The inclusion criteria for participants were: medical professionals in the orthopaedic, sports medicine or sport rehabilitation field, with or without ultrasound experience or training. Twenty raters were recruited at a European Sports Medicine symposium to rate the scans independently, in a group setting. Subsequently, a further 10 participants were recruited through opportunistic sampling (see Table 1 for characteristics). This group viewed the scans individually on a standard size desktop PC computer (screen size 50 x28 cm). These participants received the same presentation on thoracolumbar fascia. All scans were anonymised and displayed in randomised order. All participants viewed all 30 scans. Participants were asked about clinical training, years of clinical experience, musculoskeletal ultrasound training, and frequency of ultrasound image usage for diagnostic purposes in clinical practice.

Table 1: Characteristics of raters

Ultrasound image data acquisition

Images were taken at the intervertebral level 2-3, as fascial planes are the most parallel to the skin at this level [4]. The interspinous ligament between lumbar vertebrae 2 and 3, and the superficial border of posterior paraspinal muscles were identified using a validated protocol [21]. One focal region was set as close as possible to the thoracolumbar complex. Bi-lateral parasagittal (longitudinal) images were taken 2 cm lateral of the intervertebral disc space between lumbar vertebrae 2 and 3. The image acquisition was based on a validated protocol [4]. All images presented to raters were obtained using uniform settings, a frequency of 18MHz was used, with a depth of 3 cm, which allow optimum image quality for subcutaneous structures [22]. See Figure 1 for example of ultrasound image and anatomical orientation.

Each ultrasound image was obtained using B-Mode imaging, with a MyLabGold25 semi-portable ultrasound scanner (Easote, Rimini, Italy). A 4 cm, 18MHz linear array transducer (Easote LA435) was used for all images.

Selection of ultrasound images for reliability study

Initially, a single investigator selected 40 scans from a data-base of 308 bi-lateral scans of 154 male and female subjects with and without lower back pain from a larger prior study. A focus group then viewed the 40 images and selected 30 scans. Both the individual investigator and the focus group were instructed to select scans which, in their opinion, represented both ‘organised’ perimuscular fascia and ‘disorganised’ perimuscular fascia, with a range in between. ‘Organised’ was defined as ‘being able to draw a rectangular box’ around the hyperechoic zone, ‘disorganised’ was described as ‘not being able to draw a rectangular box’ around the hyperechoic zone. All raters were blind to any pathology or background information related to the scans. These 30 scans were deemed to represent the range of morphologies from very disorganised to very organised and a range of scans in between (Figure 2).

Inter-observer reliability rating protocol

In inter-observer reliability studies, it is vital that raters apply coding to data they understand [23]. For this reason, a 20 mins presentation about the thoracolumbar fascia was delivered, this facilitated anatomical orientation and exposed the participants to a representative range of ultrasound images prior to rating. Participants were not given examples of actual ratings, only of the range of images they would be rating, to avoid bias. (See Figure 1 for anatomical orientation and region of interest). Scans were projected on a standard sized screen (133 x 100 cm).

Table 1 shows that 57% had no training or experience in ultrasound imaging, 40% had experience ranging from monthly to daily evaluations of ultrasound imaging, 1 participant did not respond to this question, no observers had experience in evaluating ultrasound images of thoracolumbar fascia.

Participants were instructed to rank the region of interest (ROI in Figure 1) which included the thoracolumbar fascia (* thoracolumbar fascia in Figure 1) and the subcutaneous zone (*SZ in Figure 1) on a Likert-type scale. A Likert scale with rating points from 1 to 10 was used, point 1 was labelled as 'very disorganised' and point 10 as 'very organised', the intermediate points were numbered but remained unlabelled. Participants were familiarised to the definition of thoracolumbar fascia organisation and disorganisation. For instance, 'very organised' was defined as 'to be able to draw a rectangular shaped box around the hyperechoic area of thoracolumbar fascia' (see Figure 1).

Participants viewed scans sequentially in a time frame of 30 seconds to 1 minute. They were able to modify responses, request to re-assess a scan, and make written comments about their decisions. Participants could not discuss ratings with each other, in order to avoid bias. All responses were anonymised prior to analysis.

Statistical analysis

Inter-rater reliability was assessed from the total raw scores of all 899 decisions, and the raw scores divided into 4 sub-groups using Cronbach's alpha, to assess internal consistency among observers [24, 25]. The Cronbach's alpha coefficient was calculated using SPSS (version 21) statistical software. Standard error of measurement (SEM) was calculated as the square root of error variance in accordance with de Vet's guidelines [26]. The Krippendorff's alpha for ordinal measures was used to assess inter-observer agreement [23, 27] and was calculated using a custom-designed online calculator [28]. As Likert scales are an ordinal measurement, the median and interquartile range for the total of scans was calculated, as well as for each scan individually [29, 30].

Participant ratings of scans were categorised into four groups [30–32]. Group 1 (very disorganised) consisted of all scans with a median rating of 1 to 3. Group 2 (somewhat disorganised) consisted of all median ratings from 4 to 5. Group 3 (somewhat organised) consisted of all median ratings from 6 to 7. Group 4 (very organised) consisted of all median ratings from 8 to 10 (Figure 2). The Cronbach's alpha and Krippendorf's alpha were calculated using the original raw scores from individual raters for each scan.

Results

Results of descriptive analysis

The median (m= 5) and interquartile range (IQR=4) of the total ratings were calculated (range = 1-10), as well as for each group (Table 2 and Figure 3).

Results of inter-rater reliability analysis

All participants assessed all scans, except one participant who did not complete one rating. The Cronbach's alpha was 0.98, which is considered excellent according to the Landis and Koch criteria [33]. Observers without ultrasound imaging experience scored a Cronbach's alpha = 0.96, observers with ultrasound imaging experience scored a Cronbach's alpha = 0.95, both in the excellent range. Scores between 4 sub-groups are reported in Table 2. The Krippendorff's alpha for ordinal measures was .61, with an error variance of 0.63, indicating a modest degree of agreement.

Table 2: Inter-rater reliability scores for all data and sub-groups

Discussion

In this study we found that medical practitioners agree on different morphological features in ultrasound images of thoracolumbar fascia such as levels of organisation and disorganisation. This agreement is independent of experience in ultrasound image rating. We found that the knowledge gap between musculoskeletal (MSK)-trained radiologists, MSK-trained medical doctors and physiotherapists on the one hand, and clinicians untrained and inexperienced in MSK ultrasound, did not affect the inter-observer agreement.

It is important to establish internal consistency before images can be used for research or clinical evaluation to ensure validity [24]. The measurement error was smaller in both groups of disorganised scans, and higher in the more organised groups. This could be an indication that it may be easier to interpret disorganisation or irregular shapes rather than organisation or regular shapes. The modest Krippendorf's alpha for the ratings suggests that a minimal amount of measurement error was introduced by the independent observers, and therefore statistical power for subsequent analyses is not substantially reduced.

In this cohort, the differences in ultrasound experience do not appear to impact on consistency. We did not observe any raters who systematically under- or over-rated the images. Novice raters have demonstrated good to excellent reliability in measuring abdominal and lumbar muscle thickness obtained by ultrasound scans [34, 35]. However, a straightforward comparison between quantitative measures of lumbar and abdominal muscle tissue, commonly found in the literature on rehabilitation of lower back pain, and this study's qualitative ratings of subcutaneous connective tissue requires caution. Substantial observer variability can occur, even at the expert level of image interpretation [36]. Interestingly, in this study, experienced radiologists agreed with the interpretation of clinicians relatively inexperienced in the reading of ultrasound images. The American College of Radiology Imaging Network (ACRIN) has highlighted that in order to improve the research in interpretation of medical images, observers in reliability studies should ideally reflect a broad range of experience to provide a sufficient level of generalisability [37].

In multi-reader medical image interpretation, the phenomenon of 'groupthink', has been identified, where the opinion of novice raters might be influenced by senior or experienced raters [36]. In order to avoid a situation of potential pseudo-consensus, all raters viewed the scans independently without discussing decisions with each other.

This study has a number of limitations. First, it involved a small cohort size of both observers and scans. The results are encouraging and should be validated in a larger cohort [37]. Secondly, the study relied on static ultrasound images. Future studies may consider functional and dynamic measurements. Finally, we did not determine the frequency in which raters interpret the same image differently. This needs to be taken into account for future studies.

Conclusion

Medical practitioners agree on morphological features such as levels of organisation and disorganisation in ultrasound images of thoracolumbar fascia, regardless of experience. These findings will be useful for the establishment of a clinical diagnostic scale and the further development of using ultrasound as a decision-making tool for researchers and clinicians.

Declarations

Ethics approval and consent to participate:

This study was approved by the University of Kent’s Research and Ethics Committee (Prop. 163 – 2013). Informed consent was obtained from all participants.

Consent for publication:

Consent for publication was sought from all participants whose images are contained in this manuscript.

Availability of data and materials:

The datasets analysed during the current study are available in the KAR repository, https://kar.kent.ac.uk/view/

Competing interests:

The authors declare they have no competing interests.

Funding:

Not applicable

Authors' contributions:

KDC conceived the study, participated in study design, collected the data, analysed the data and drafted the manuscript

KH participated in study concept and design, reviewed the manuscript

JWD participated in study design, analysis, interpretation and manuscript preparation

LP participated in study design and manuscript preparation

All authors read and approved the final manuscript.

Acknowledgements:

The authors thank Karthik Muthumayandi for assistance with development of electronic data collection, and Dr Samantha L Winter for helpful discussions during manuscript preparation.

References:

1. Langevin HM, Sherman KJ: **Pathophysiological model for chronic low back pain integrating connective tissue and nervous system mechanisms.** *Med Hypotheses* 2007, **68**:74–80.

2. Yahia LH, Rhalmi S, Newman N, Isle M: **Sensory innervation of human thoracolumbar fascia**. *Acta Orthop Scand* 1992, **63**:195–197.

3. Taguchi T, Tesarz J, Mense S: **The Thoracolumbar Fascia as a Source of Low Back Pain**. In *Fascia Research II - Basic Science and Implications for Conventional and Complementary Healthcare*. Edited by Huijing PA, Hollander P, Findley TW, Shleip R. Munich: Elsevier GmbH; 2009:251.

4. Langevin HM, Stevens-Tuttle D, Fox JR, Badger GJ, Bouffard N a, Krag MH, Wu J, Henry SM: **Ultrasound evidence of altered lumbar connective tissue structure in human subjects with chronic low back pain.** *BMC Musculoskelet Disord* 2009, **10**:151.

5. Tesarz J, Hoheisel U, Wiedenhöfer B, Mense S: **Sensory innervation of the thoracolumbar fascia in rats and humans.** *Neuroscience* 2011, **194**:302–8.

6. Hoheisel U, Taguchi T, Treede RD, Mense S: **Nociceptive input from the rat thoracolumbar fascia to lumbar dorsal horn neurones**. *Eur J Pain* 2011, **15**:810–815.

7. Langevin HM, Fox JR, Koptiuch C, Badger GJ, Greenan-Naumann AC, Bouffard N a, Konofagou EE, Lee W-N, Triano JJ, Henry SM: **Reduced thoracolumbar fascia shear strain in human chronic low back pain.** *BMC Musculoskelet Disord* 2011, **12**:203.

8. Willard FH, Vleeming A, Schuenke MD, Danneels L, Schleip R: **The thoracolumbar fascia: anatomy, function and clinical considerations.** *J Anat* 2012, **221**:507–36.

9. Barker PJ, Hapuarachchi KS, Ross JA, Sambaiew E, Ranger TA, Briggs CA: **Anatomy and biomechanics of gluteus maximus and the thoracolumbar fascia at the sacroiliac joint**. *Clin Anat* 2014, **27**:234–240.

10. Macintosh JE, Bogduk N, Gracovetsky S: **The biomechanics of the thoracolumbar fascia.** *Clin Biomech* 1987, **2**:78–83.

11. Vleeming A, Pool-Goudzwaard AL, Stoeckart R, van Wingerden JP, Snijders CJ: **The posterior layer of the thoracolumbar fascia. Its function in load transfer from spine to legs.** *Spine (Phila Pa 1976)* 1995, **20**:753–758.

12. Carvalhais VO do C, Ocarino J de M, Araujo VL, Souza TR, Silva PLP, Fonseca ST: **Myofascial force transmission between the latissimus dorsi and gluteus maximus muscles: An in vivo experiment**. *J Biomech* 2013, **46**:1003–1007.

13. Li W, Ahn AC, Weitz D, Mahadevan L, Barnett R, Zhang M: **Subcutaneous Fascial Bands—A Qualitative and Morphometric Analysis**. *PLoS One* 2011, **6**:e23987.

14. Pavan PG, Stecco A, Stern R, Stecco C: **Painful connections: Densification versus fibrosis of fascia**. *Curr Pain Headache Rep* 2014, **18**:441.

15. Diviti S, Gupta N, Hooda K, Sharma K, Lo L: **Morel-lavallee lesions-review of pathophysiology, clinical findings, imaging findings and management**. *J Clin Diagnostic Res* 2017, **11**:TE01-TE04.

16. Corey SM, Vizzard M a, Bouffard N a, Badger GJ, Langevin HM: **Stretching of the back improves gait, mechanical sensitivity and connective tissue inflammation in a rodent model.** *PLoS One* 2012, **7**:e29831.

17. Schilder A, Hoheisel U, Magerl W, Benrath J, Klein T, Treede R-D: **Sensory findings after stimulation of the thoracolumbar fascia with hypertonic saline suggest its contribution to low back pain.** *Pain* 2014, **155**:222–31.

18. Bishop JH, Fox JR, Maple R, Loretan C, Badger GJ, Henry SM, Vizzard MA, Langevin HM: **Ultrasound Evaluation of the Combined Effects of Thoracolumbar Fascia Injury and Movement Restriction in a Porcine Model**. *PLoS One* 2016, **11**:e0147393.

19. Langevin HM, Fox JR, Koptiuch C, Badger GJ, Greenan- Naumann AC, Bouffard NA, Konofagou EE, Lee W-N, Triano JJ, Henry SM: **Reduced thoracolumbar fascia shear strain in human chronic low back pain**. *BMC Musculoskeletal Disorders* 2011:203.

20. Klingler, Werner, Velders M, Hoppe K, Pedro M SR: **Clinical Relevance of Fascial Tissue and dysfunctions**. *Curr Pain Headache Rep* 2014, **18**:439.

21. Stokes M, Hides MJ, Elliott J, Kiesel MSK, Hodges CP, Hons B: **Rehabilitative Ultrasound Imaging of the Posterior Paraspinal Muscles**. *J Orthop Sport Phys Ther* 2007, **37**:581–595.

22. Kremkau F: *Diagnostic Ultrasound: Principles and Instruments*. 7th Edition. St. Louis, Mo: Saunders; 2006.

23. Krippendorff K: **Reliability in content analysis: some common misconceptions and recommendations**. *Hum Commun Res* 2004, **30**:411–433.

24. Tavakol Mohsen DR: **Making sense of Cronbach's alpha**. *Int J Med Educ* 2011, **2**:53–55.

25. Cronbach LJ, Shavelson RJ: **My Current Thoughts on Coefficient Alpha and Successor Procedures**. *Educ Psychol Meas* 2004, **64**:391–418.

26. de Vet HCW, Terwee CB, Knol DL, Bouter LM: **When to use agreement versus reliability measures**. *J Clin Epidemiol* 2006, **59**:1033–1039.

27. Hayes A F KK: **Answering the call for a standard reliability measure for coding data**. *Commun Methods Meas* 2007, **1**:77–89.

28. Freelon D: **ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service**. *Int J Internet Sci* 2013, **8**:10–16.

29. Jamieson S: **Likert scales: How to (ab)use them**. *Med Educ* 2004, **38**:1217–1218.

30. Norman G: **Likert scales, levels of measurement and the “laws” of statistics**. *Adv Heal Sci Educ* 2010, **15**:625–632.

31. LaValley MP, Felson DT: **Statistical presentation and analysis of ordered categorical outcome data in rheumatology journals.** *Arthritis Rheum* 2002, **47**:255–259.

32. Hallgren KA: **Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial**. *Tutor Quant methods Psychol* 2012, **8**:23–34.

33. Landis JR, Koch GG: **The Measurement of Observer Agreement for Categorical Data**. *Biometrics* 1977, **33**:159–174.

34. Teyhen DS, George SZ, Dugan JL, Williamson J, Neilson BD, Childs JD: **Inter-rater reliability of ultrasound imaging of the trunk musculature among novice raters.** *J Ultrasound Med* 2011, **30**:347–56.

35. Wilson A, Hides JA, Blizzard L, Callisaya M, Cooper A, Srikanth VK: **Measuring ultrasound images of abdominal and lumbar multifidus muscles in older adults: A reliability study**. *Man Ther* 2016.

36. Bankier A a, Levine D, Halpern EF, Kressel HY: **Consensus interpretation in imaging research: is there a better way?** *Radiology* 2010, **257**:14–17.

37. Obuchowski NA: **How Many Observers Are Needed in Clinical Studies Peter Hogg**. *Am J Roentgenol* 2004, **182**(April):867–869.

Tables:

Table 1: Characteristics of raters

Clinical training	**N=30**
MD	21 (70%)
Physiotherapists	7 (23%)
Radiologists	2 (6%)
Years of clinical experience	13.03 (± SD 9.6)
USI training & experience	**N=30**
Trained & experienced	12 (40%)
Untrained & unexperienced	17 (57%)
not known	1 (3%)
Frequency of USI usage	**n=12 (40%)**
daily	4 (33%)
weekly	4 (33%)
monthly	4 (33%)

USI = ultrasound imaging

Table 2: Inter-rater reliability scores for all data and sub-groups

Group	**Decisions (%)**	**Median (IQR)**	**Cronbach's alpha**	**Landis and Koch criteria** [33]	**SEM**
All data	899	5 (4)	.98	excellent	0.10
Group 1	300 (32.8%)	2 (3)	.70	excellent	0.40
Group 2	209 (22.6%)	5 (3)	.68	good	0.17
Group 3	150 (20.3%)	7 (3)	.47	moderate	0.56
Group 4	240 (24.2%)	8 (2)	.56	moderate	0.50

SEM = standard error of measurement. Group 1 = very disorganised. Group 2 = somewhat disorganised. Group 3 = somewhat organised. Group 4 = very organised.

List of Figures with legends:

Figure 1: Anatomical orientation and delineation of the zones rated.

Figure 1. Anatomical orientation and delineation of the zones rated.
*D = dermis. *SZ = subcuteanous zone. *TFL = thoracolumbar fascia.
*ES = erector spinae. ROI = region of interest, zones rated.

Figure 2: A range of different thoracolumbar fascia morphologies.

Figure 2. Sub-groups of different TLF morphologies. Group 1 =example of 'very disorganised', Group 2= 'somewhat disorganised' Group 3= 'somewhat organised', Group 4= 'very organised' . The sub-grouping was based on the median scores for each scan.

Figure 3: Box-plots of all ratings, and ratings for each sub-group.

Figure 3. Boxplots for total scores of the ratings (899 decisions) and ratings for each sub-group. Central tendency is the median, distribution is the interquartile range.