Original Research

ULTRASOUND

Ultrasound
2016, Vol. 24(3) 134–141

DOI: 10.1177/1742271X16652439
ult.sagepub.com

SAGE

Development and assessment of inter- and intra-rater reliability of a novel ultrasound tool for scoring tendon and sheath disease – A pilot study

L Horton[1,2], P Emery[1,3] and P Marshall[2]

Abstract

Background: Tendon and synovial sheath disease is common. A method of monitoring the status of tendons and sheaths is important for both diagnosis of pathology and evaluation of the efficacy of treatments. For this study, an ultrasound scoring tool was developed and its reliability tested between raters. The tool is novel in that it scores tendons and sheaths separately, an important consideration since disorders of these structures are not necessarily concurrent.

Methods: Thirty diseased tendons and sheaths were included in this pilot cross-sectional study. Tendon and sheath measurements were taken and the semi-quantitative five-grade score was applied to assess tendon greyscale, tendon Doppler activity and sheath Doppler activity. Inter-rater and intra-rater agreement exercises were undertaken to test the reliability of the scoring tool.

Results: The Intra-class Correlation Coefficient values for both the inter-rater and intra-rater reliability tests showed excellent agreement for the tendon and sheath measurements. Unweighted kappa estimations for inter-rater scores showed excellent agreement for tendon Doppler; good agreement was shown for scoring sheath Doppler, while poor agreement was shown for tendon grey-scale scoring. The intra-rater reliability scores demonstrated similar results.

Conclusion: Overall, the study strongly supports the use of this scoring tool for the diagnosis and follow-up of tendon and sheath disorders. The results may be used as a starting point from which to base further work in this important area. Future studies should address the limitations found in this research with a strong focus on improving tendon grey-scale measurement accuracy and agreement.

Keywords

Tendon degeneration, ultrasound scoring tool, tendon sheath

Date received: 18 February 2016; accepted: 28 April 2016

Introduction

It is widely accepted that tendons can be recalcitrant to treatment, highlighting the need for knowledge acquisition through the study of tendon disorders as well as assessment of the efficacy of the various therapies.[1] Tendons and tendon sheaths can be affected by acute or overuse injuries, age-related degeneration and inflammatory conditions such as rheumatoid arthritis.[2] Matrix derangement, which occurs as a result of injury and disease, places tendons at increased risk of tearing

[1]NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds, UK
[2]School of Healthcare, Baines Wing, University of Leeds, Leeds, UK
[3]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

Corresponding author:
L Horton, NIHR Leeds Musculoskeletal Biomedical Research Unit, Chapel Allerton Hospital, Leeds, LS7 4SA UK.
Email: laura.horton1@nhs.net

and rupture. Tendon rupture is rarely an acute event and is usually the result of pre-existent degeneration,[2–5] which may have been clinically silent up until the time of the rupture.[6] It has been estimated that 97% of tendons that rupture have degenerative changes,[1] and this has been confirmed at surgical repair where the remainder of the tendon tissue is found to be tendinopathic.[7,8] Inflammatory and degenerative models of tendinopathy have traditionally been described, but the absence of inflammatory cells at histology plus a lack of evidence to support a theory of non-reparative disease has resulted in the proposal of an alternative 'continuum' model of tendon pathology.[9,10] This model which places a tendon's pathological status anywhere within three overlapping phases of matrix change was developed and described[1] and subsequently cited in several studies.[11–13] It is an important histological and clinical progression from the long-established view of tendon disease and treatment, as it outlines the stages at which reversal of pathology is possible and at what point there is little chance of repair. In order to utilise this model in clinical and research practice, a method of monitoring the matrical health status of tendons and sheaths is required to provide baseline diagnostic scores of disease as well as follow-up evaluations to investigate efficacy of treatments. Currently, there are no clear guidelines for the diagnosis and management of tendon and sheath disorders, and clarity is required regarding the interventions and techniques used for facilitating the process of repair.

Ultrasound is a non-invasive, accessible and rapid imaging method which shows high reliability in the assessment of tendon abnormalities.[14,15] Moreover, continuing technological developments resulting in extremely high-resolution images, coupled with the real-time capabilities of ultrasound, enable accurate diagnoses of soft-tissue disorders.[16,17] The aim of this pilot study was to test a novel ultrasound scoring tool developed for the purpose of grading disorders of tendons and sheaths separately, a necessary approach, since disease in these two structures is not necessarily concurrent. The histological composition of normal and diseased tendons relates well to the ultrasound appearances of these structures, given that the pathological process causes a change in morphology, and therefore, altered acoustic interfaces within the tendon matrix.[1,6] Some histological and morphological changes that occur in diseased tendons and sheaths and which result in specific ultrasound appearances are considered to be significant and measurable using ultrasound technology (see Table 1).[18]

Other abnormalities such as intra-sheath effusions, tendon tears and calcific deposits, although relevant parameters in evaluation of tendon health, may be suitable for interventional techniques, so require full description in terms of size, location, chronicity and in the case of tears, fibre retraction measurements.

Table 1. Measurable characteristics of tendon degeneration and tenosynovitis

Measurable characteristics of tendon degeneration and tenosynovitis
• Maximum tendon thickness
• Amount of altered/decreased echogenicity and separation of tendon fibres
• Amount of Doppler flow in tendon
• Sheath thickness
• Sheath Doppler flow

Consideration of the characteristics in Table 1 resulted in the development of the ultrasound scoring tool. The simple parameters and ease of use of a previously reported sheath Power Doppler (PD) grading method[19] led to its adaptation and modification for the current study. However, the tendon greyscale (GS) and PD scoring method was innovative.

Methods

Ethical approval was gained from NRES Committee Yorkshire & The Humber – Bradford Leeds, Ref 14/YH/1279, and Leeds Teaching Hospitals Trust Research and Innovation, R&I Number RR14/11381.

A sample of 16 patients from rheumatology clinics, who were found at their routine clinical scan to have ultrasound identified ankle or wrist tendon and sheath disease, were recruited to the study. Adults over 18 years of age who were able to give informed consent were included. A cohort of adults was required to represent the target population, since human anatomical musculoskeletal development has not reached maturity until 18 years of age, therefore, the appearances of tendons and their bony attachments can differ in children and young adults. The only exclusion criterion was that the potential participant was unable to provide informed consent.

All informed participants elected to consent and have the research scan completed on the same day as their clinic appointment, thereby preventing the need for a separate trip to the research unit on a different day. The lead research sonographer performed both the initial clinic scan and the subsequent research scan. Since this was a pilot study, a sample size adequate for producing a range of pathology grades was required. Wrist and ankle tendons with synovial sheaths were targeted for this pilot study as they are most commonly seen in this department. The extensor carpi ulnaris (ECU) and extensor and flexor digitorum tendons were evaluated in

the wrist and the tibialis posterior and peroneal tendons in the ankle. Since the tool scores tendons and sheaths separately, it was unnecessary to include non-synovial tendons such as the Achilles. Aetiology was mixed, with some patients reporting mechanical injury and others, whose symptoms coincided with the onset of inflammatory disease. For the purposes of this study, no distinction was made between the two groups. Some patients had more than one abnormal structure and from the 16 patients recruited, 30 tendons and sheaths were found to have varying grades of disease, which were then included in the study.

A second group of four healthy controls was recruited in order to demonstrate zero pathology. The smaller control group was considered sufficient as there can only be one true representation of normal. Asymptomatic adults 30 years of age or less with no history of wrist or ankle trauma or pain were recruited. Tendons in this age group are less likely to have age-related degenerative changes. The purpose of the control group was to demonstrate normal morphology of tendons and sheaths as they are described in the literature and to provide a base comparison on which to use the scoring tool. Written informed consent was obtained from all participants.

The ultrasound scans were carried out by one experienced musculoskeletal sonographer (LH) using a Logiq E9 (General Electric (GE) Healthcare, Chalfont Saint Giles, UK) equipped with a multi-frequency (5–16 MHz) linear transducer and software version R5, Revision 1.1. The B-mode and PD settings were optimised for maximal image resolution and flow sensitivity in superficial soft-tissue structures as follows:

B-mode: Frequency = 15 MHz, Gain = 52.
PD mode: Frequency = 10 MHz, Pulse Repetition Frequency = 0.8 kHz, Gain = 15.

Other than to adjust the scanning depth, these settings were not changed throughout the study.

The ultrasound examination was carried out firstly to confirm the presence of tendon and/or sheath disease at the region of pain indicated by the patient. B-mode ultrasound was performed in order to assess tendon GS echogenicity and tendon and sheath thickness (Figure 1). PD ultrasound was used to detect intratendinous vascularity as well as tenosynovitis. If pathology was proven at the scan, one transverse image was captured at the site of maximal disease focus, and this image was used to apply measurements according to the pre-determined scoring tool. Normal tendons and sheaths demonstrate avascularity at ultrasound scan, therefore, there is scarcity of Doppler signals in healthy structures.[20] The reference for assessing intratendinous GS levels was the echogenic tendon collagen fibres.

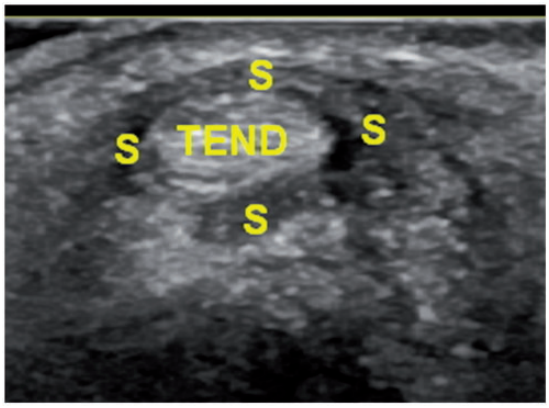

Figure 1. Cross-sectional ultrasound image of normal extensor carpi ulnaris (ECU) tendon (Tend) encircled by the synovial sheath (S).

Comparison of the bright echoes from these fibres was made with the relatively darker tissue which is present in varying amounts depending on the absence or presence and degree of tendon pathology.

The inter-rater and intra-rater reliability exercises were carried out by two experienced musculoskeletal sonographers between three and six weeks after the final participant had been scanned and required the recall of the original saved images from the participants' imaging record on the ultrasound unit archive. Once displayed on the screen, the raters applied the same measurement and Doppler scoring procedure as that undertaken at the actual scan. Both sonographers were blinded to the original measurements.

Tendon measurements

Using the single captured transverse image, the tendon was measured in the region displaying maximum disease level, as characterised by tendon thickening due to collagen fibre separation and increased ground substance. Any associated sheath thickening and/or inflammation was also measured and scored. This was achieved by placement of electronic measurement callipers on the surface-to-depth leading-edge opposing borders of the tendon or sheath and measurements taken to 0.1 mm. One tendon measurement and one sheath measurement were acquired at each site, an approach taken in an attempt to keep the process as simple as possible (Figure 2). In order for a scoring tool to be useful in practice, its use should be uncomplicated.[21]

The GS characteristics of the tendon were subjectively assessed and assigned a semi-quantitative score (0–4) where:

0 = normal echotexture
1 = up to 25% of affected area of tendon shows GS pathological change

2 = up to 50% of affected area of tendon shows GS pathological change
3 = up to 75% of affected area of tendon shows GS pathological change
4 = up to 100% of affected area of tendon shows GS pathological change

Doppler flow signals were subjectively assessed (Figure3) and assigned a semi-quantitative score (0–4) where:

0 = no abnormal Doppler
1 = up to 25% of degenerated area shows PD signal
2 = up to 50% of degenerated area shows PD signal
3 = up to 75% of degenerated area shows PD signal
4 = up to 100% of degenerated area shows PD signal

Sheath measurements

Electronic measurement callipers were placed on the sheath borders at the region of greatest thickness and measurements taken to 0.1 mm (Figure 4).

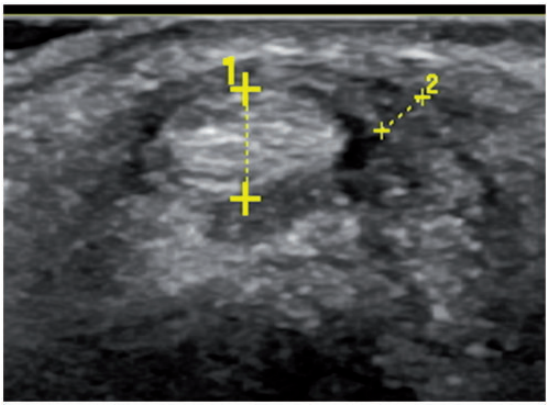

Figure 2. Electronic measurement calliper placement to obtain diameter of tendon and thickness of sheath.

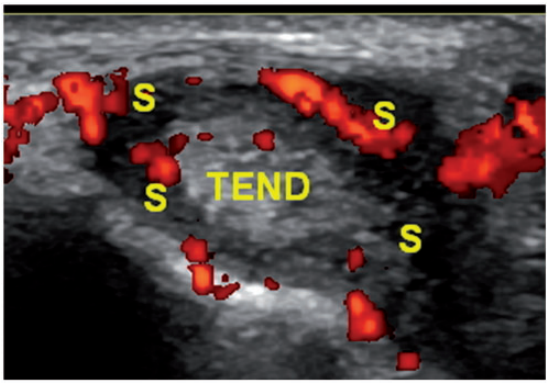

Figure 3. Cross-sectional image of tendon (Tend) and sheath showing Doppler flow within tendon and thickened sheath (S).

Doppler flow signals were subjectively assessed and allocated a semi-quantitative score (0–4) where:

0 = no abnormal Doppler
1 = up to 25% of pathological area shows PD signal
2 = up to 50% of pathological area shows PD signal
3 = up to 75% of pathological area shows PD signal
4 = up to 100% of pathological area shows PD signal

Figures 5 to 7 show examples using the novel scoring method, where the letter T prefixes the tendon score of thickness in mm, GS score and PD score. The letter S prefixes the sheath scores of thickness and PD.

In the first example (Figure 5), the ECU tendon has a maximum diameter of 3.3 mm, between 75% and 100% of the affected tendon shows greyscale matrix change, and there is Doppler signal in less than 25% of the pathological portion of tendon. The sheath maximum thickness is 1 mm, and no Doppler signal is present.

These data produce a score of T3.3, 4, 1/S1.0, 0

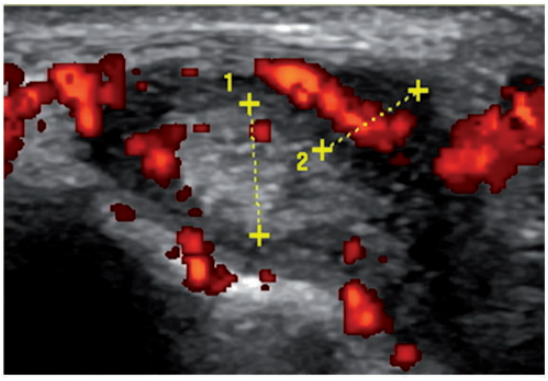

Figure 4. Electronic measurement calliper placement to quantify greatest tendon diameter and thickest region of sheath.

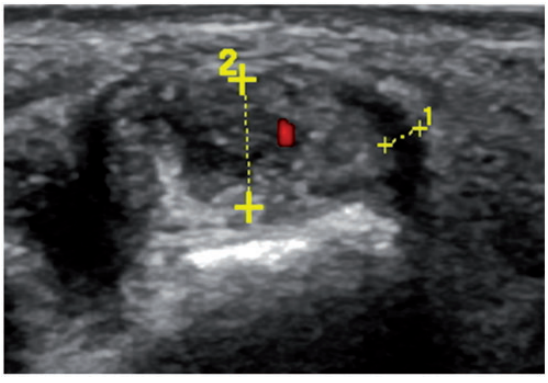

Figure 5. **T**3.3, 4, 1/**S**1.0, 0.

In Figure 6, the ECU tendon has a maximum diameter of 3.5 mm, up to 25% of the affected tendon shows greyscale matrix change, and there is Doppler signal in up to 25% of the pathological portion of tendon. The sheath maximum thickness is 2.1 mm, and Doppler signal is present in 25% to 50% of the thickened sheath.

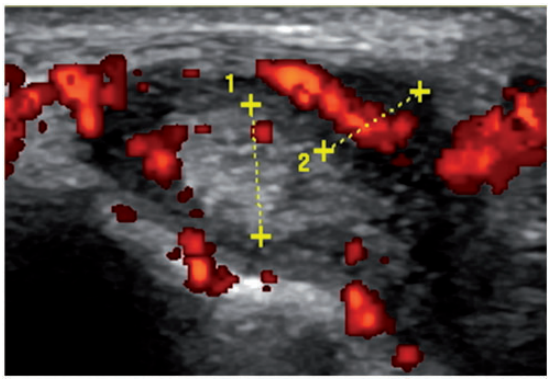

Figure 6. **T**3.5, 1, 1/**S**2.1, 2.

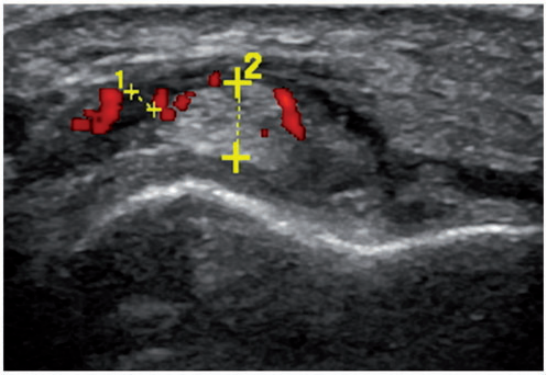

Figure 7. **T**2.7, 0, 1/**S**1.1, 1.

These data produce a score of T3.5, 1, 1/S2.1, 2

In Figure 7, the ECU tendon has a maximum diameter of 2.7 mm, zero greyscale matrix change, and there is Doppler signal in up to 25% of the pathological portion of tendon. The sheath maximum thickness is 1.1 mm, and Doppler signal is present in up to 25% of the thickened sheath.

These data produce a score of T2.7, 0, 1/S1.1, 1

Results

The data were analysed using SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp). All data created during this research are available on request from the corresponding author.

Continuous data: Tendon and sheath measurements (mm)

Inter-rater agreement was estimated using calculation of intra-class correlation coefficients (ICC) (two-way mixed effects model, consistency definition). Intra-rater agreement was estimated using calculation of ICC (one-way random model, consistency definition). ICCs were interpreted according to Portney and Watkins (Table 2), who state that values below 0.5 can be thought of as representing 'poor reliability'. Values between 0.51 and 0.75 indicate 'moderate reliability' and values above 0.75 represent 'good reliability'.[22]

Bland Altman plots (Table 2) were also produced to explore the relationship between results in the reliability exercises. Inter-rater and intra-rater differences were estimated by calculating the mean (and 95% confidence

Table 2. ICC and Bland Altman results of reliability tests

	ICC			Bland Altman	
	Value	95% CI	Bias	95% Limits of agreement	SD
Inter-rater					
Tendon	0.94	(0.88, 0.97)	0.06	(−0.78, 0.66)	0.37
Sheath	0.80	(0.62, 0.90)	0.21	(−0.86, 0.43)	0.33
Intra-rater					
Tendon	0.95	(0.89, 0.97)	0.03	(−0.69, 0.75)	0.37
Sheath	0.81	(0.64, 0.90)	0.12	(−0.74, 0.50)	0.32

ICC: intra-class correlation coefficients; SD: standard deviation; CI: confidence intervals.

Table 3. Inter-rater reliability kappa estimations with 95% CI

Inter-rater reliability kappa estimations with 95% CI		
Tendon GS	K = 0.18	(95% CI: − 0.07, 0.43)
Tendon PD	K = 0.81	(95% CI: 0.68, 0.94)
Sheath PD	K = 0.64	(95% CI: 0.39, 0.89)

PD: power Doppler; GS: greyscale; CI: confidence intervals.

Table 4. Intra-rater reliability kappa estimations with 95% CI

Intra-rater reliability kappa estimations with 95% CI		
Tendon GS	K = 0.08	(95% CI: − 0.1, 0.27)
Tendon PD	K = 0.87	(95% CI: 0.57, 1.0)
Sheath PD	K = 0.68	(95% CI: 0.43, 0.93)

PD: power Doppler; GS: greyscale; CI: confidence intervals.

interval (CI)) of the differences between measurements generated in the reliability tests. Variability was calculated as 1.96 × standard deviation (SD) of these differences. This gives the estimated bias value and should be as close to zero as possible. Normal distribution will place 95% of the differences within the upper and lower limits of agreement, a range of ± 1.96 × SD of the mean difference.[23] The ICC and Bland Altman results are presented in a side by side format (see Table 2) to facilitate comparison of results between analysis methods. The ICC values for both the inter-rater and intra-rater reliability tests showed excellent agreement for the tendon and sheath measurements.

The Bland Altman plots showed normal distributions of the mean differences with no proportional bias seen in the data placement above and below the bias line. The small bias values and limits of agreement were clinically acceptable and indicative of the lack of systematic error in the tests. This level of agreement was present over the range of values.

Categorical data: Five-grade semi-quantitative scale

Unweighted kappas (which only assess presence or absence of absolute agreement) were calculated with 95% CIs. Kappa statistics were interpreted according to the published guidelines, where 0.01–0.20 = poor agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = good agreement and 0.81–0.99 = excellent agreement.[24]

In order to reflect the sampling error present, 95% CI were calculated for each kappa value using the formula: $K - 1.96 \times SE$ to $K + 1.96 \times SE$. Table 3 shows the unweighted kappa estimations for inter-rater scores, where tendon PD showed excellent agreement. Good agreement was shown for scores of sheath PD, while poor agreement was shown for tendon GS scoring.

The intra-rater reliability scores demonstrated similar results with excellent agreement for tendon PD scores. Good agreement was shown for scores of sheath PD, while poor agreement was shown for tendon GS scoring (Table 4).

Discussion

To date, several sheath disease scoring tools have been tested for use in rheumatology.[25–29] However, no papers have been published which specifically tested a method to quantify tendon degeneration. Previous studies have concentrated on quantification only of the sheath PD.[25–29] However, Bruyn et al.[27] carried out a large study which aimed to develop the first ultrasound scoring system of tendon damage in rheumatoid arthritis. A semi-quantitative 0–2 score was specified, where 0 = normal, 1 = partial tear and 2 = complete rupture. This information would be of limited clinical use, since partial tearing or rupture is the end-stage of the degeneration continuum and not amenable to matrix regeneration. This pilot study is the first to separately assess tendon and sheath disease and to test the feasibility of a disease scoring tool. The two structures are closely associated in that they are two components of an anatomical unit; however, they do not automatically behave as a single structure pathologically. In cases of mechanical injury, tendon derangement does not necessarily correlate with or lead to sheath pathology. Current research highlights the continuum nature of the tendon disease process, with healing and reversal of matrix breakdown shown to occur with appropriate and timely treatment. In cases of rheumatological disease, a normal appearing tendon may be enveloped by a severely diseased sheath or, the two structures may show characteristics of concurrent disease.[28]

The excellent agreement shown in this study reflects previously reported results,[30] which used a broader 0–3 semi-quantitative scale. The kappa results support the inter-rater and intra-rater reliability of the semi-quantitative scale for tendon and sheath PD with the agreement results favourable compared with other studies.[28,31]

The current study does highlight the poor reliability of the tendon GS scoring method, which is almost certainly due to the difficulty of retrospectively

interpreting a single static ultrasound image. Without the ability to assess a structure in several anatomical planes with the ultrasound transducer angled for optimal visualisation of tissue borders and matrix characteristics, diagnostic accuracy is difficult to achieve. This could be rectified by initially storing more images, or incorporating an assessment component whereby the inter- and intra-rater reliability methods are analysed in order to identify and reduce measurement error. Operator dependency is also a well-documented limitation of ultrasound as a diagnostic technique, and this applies to both acquisition and interpretation of the images.[16] One of the aims of a scoring tool should be to reduce systematic error caused by multiple subjective assessments. Conventionally, the goal and determinant of success of conservative therapies for tendon disorders has been resolution of patient reported symptoms and clinician assessed return to function. Since pain resolution cannot reliably be correlated with the reversal of the degenerative process or a return to normal tendon morphology,[32,33] its use as a measure of treatment success may not be a good predictor of clinical outcome.

Conclusion

This is the first application of a clinically relevant novel scoring method to assess and separately grade tendon and sheath disease, the promising measurement and PD results suggesting that this work may serve as a base for other projects. A reliable and validated tendon and sheath disease scoring method would inform clinicians who are involved in both short- and long-term follow-up of this debilitating disorder. It would also enable longitudinal research designed to investigate temporal changes in tendon matrical health. Future research in this area would be beneficial in order to address the poor agreement shown in the current study for tendon GS scoring, the aim of which should be to develop and evaluate a reliable measurement method for this parameter. Validation of the scoring tool should follow, with methods to include a larger sample in order to ensure all grades of disease are represented and to reduce standard error in the data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This report is independent research supported by the National Institute for Health Research/Chief Nursing Officer Clinical Academic Training Programme Masters in Clinical Research Scheme. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Ethical approval

Ethical approval was gained from NRES Committee Yorkshire & The Humber – Bradford Leeds, Ref 14/YH/1279, and Leeds Teaching Hospitals Trust Research and Innovation, R&I Number RR14/11381. Written informed consent was obtained from all participants.

Guarantor

LH

Contributorship

LH researched the literature and conceived the study. PM was involved in protocol development and gaining ethical approval. PE was involved in protocol development and patient recruitment. LH wrote the first draft and final version of the manuscript. All authors reviewed the final version of the manuscript.

Acknowledgements

The authors express gratitude to the following radiologists for their advice: Dr David Kerr, Dr Jeremy Sharr and Dr Tony Goh.

References

1. Cook JL and Purdam CR. Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *Br J Sports Med* 2009; 43: 409–416.
2. Sharma P and Maffulli N. Biology of tendon injury: healing, modeling and remodeling. *J Musculoskelet Neuronal Interact* 2006; 6: 181.
3. Jarvinen M, Jozsa L, Kannus P, et al. Histopathological findings in chronic tendon disorders. *Scand J Med Sci Sports* 1997; 7: 86–95.
4. Kannus P and Natri A. Etiology and pathophysiology of tendon ruptures in sports. *Scand J Med Sci Sports* 1997; 7: 107–112.
5. Riley G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology* 2004; 43: 131–142.
6. Rees JD, Wilson AM and Wolman RL. Current concepts in the management of tendon disorders. *Rheumatology* 2006; 45: 508–521.
7. Jarvinen TLN, Jarvinen TAH, Jarvinen M, et al. Collagen fibres of the spontaneously ruptured human tendons display decreased thickness and crimp angle. *J Orthopaed Res* 2004; 22: 1303–1309.
8. Jozsa L and Kannus P. Histopathological findings in spontaneous tendon ruptures. *Scand J Med Sci Sports* 1997; 7: 113–118.
9. Khan KMC, Jill L, Bonar F, et al. Histopathology of common tendinopathies. *Sports Med* 1999; 27: 393–408.

10. Rees JD, Stride M and Scott A. Tendons – time to revisit inflammation. *Br J Sports Med* 2014; 48: 1553.
11. Lewis JS. Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *Br J Sports Med* 2010; 44: 918–923.
12. Malliaras P and Cook J. Changes in anteroposterior patellar tendon diameter support a continuum of pathological changes. *Br J Sports Med* 2011; 45: 1048–1051.
13. McCreesh K and Lewis J. Continuum model of tendon pathology – where are we now? *Int J Exp Pathol* 2013; 94: 242–247.
14. Stevic R and Dodic M. Ultrasonography of tendon abnormalities. *OA Musculoskel Med* 2013; 1: 12.
15. Micu MC, Serra S, Fodor D, et al. Inter-observer reliability of ultrasound detection of tendon abnormalities at the wrist and ankle in patients with rheumatoid arthritis. *Rheumatology* 2011; 50: 1120–1124.
16. Patil P and Dasgupta B. Role of diagnostic ultrasound in the assessment of musculoskeletal diseases. *Ther Adv Musculoskel Dis* 2012; 4: 341–355.
17. Lee KS. Musculoskeletal sonography of the tendon. *J Ultrasound Med* 2012; 31: 1879–1884.
18. Cook J and Purdam C. Explaining ultrasound images of tendon pathology: a pathology model of load-induced tendinopathy. *Sound Eff* 2011; 16–20.
19. Alcalde M, D'Agostino MA, Bruyn GAW, et al. A systematic literature review of US definitions, scoring systems and validity according to the OMERACT filter for tendon lesion in RA and other inflammatory joint diseases. *Rheumatology* 2012; 51: 1246–1260.
20. Hodgson RJ, O'Connor PJ and Grainger AJ. Tendon and ligament imaging. *Br J Radiol* 2012; 85: 1157–1172.
21. Ohrndorf S and Backhaus M. Advances in sonographic scoring of rheumatoid arthritis. *Ann Rheum Dis* 2013; 72: ii69–75.
22. Portney LG and Watkins MP. *Foundations of Clinical Research: application to practice*, 3rd ed. New Jersey: Prentice Hall Health.
23. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
24. Landis JR and Koch GG. The Measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
25. Albrecht K, Muller-Ladner U and Strunk J. Quantification of the synovial perfusion in rheumatoid arthritis using Doppler ultrasonography. *Clin Exp Rheum* 2007; 25: 630–638.
26. Breidahl WH, Stafford Johnson DB, Newman JS, et al. Power Doppler sonography in tenosynovitis: significance of the peritendinous hypoechoic rim. *J Ultras Med* 1998; 17: 103–107.
27. Bruyn GAW, Hanova P, Iagnocco A, et al. Results of a reliability exercise for the grading of tendon rupture in patients with rheumatoid arthritis, using a consensus-based ultrasound score. *Arthritis Rheum* 2013; 65: S828.
28. Naredo E, D'Agostino MA, Wakefield RJ, et al. Reliability of a consensus-based ultrasound score for tenosynovitis in rheumatoid arthritis. *Ann Rheum Dis* 2013; 72: 1328–1334.
29. Wakefield RJ, O'Connor PJ, Conaghan PG, et al. Finger tendon disease in untreated early rheumatoid arthritis: a comparison of ultrasound and magnetic resonance imaging. *Arthritis Rheum* 2007; 57: 1158–1164.
30. Milosavljevic J, Lindqvist U, Elvin A, et al. Ultrasound and power Doppler evaluation of the hand and wrist in patients with psoriatic arthritis. *Acta Radiologica* 2005; 46: 374–385.
31. Iagnocco A, Naredo E, Wakefield R, et al. Responsiveness in rheumatoid arthritis. A report from the OMERACT 11 ultrasound workshop. *J Rheumatol* 2014; 41: 379–382.
32. Gotoh M, Hamada K, Yamakawa H, et al. Increased substance P in subacromial bursa and shoulder pain in rotator cuff diseases. *J Orthopaed Res* 1998; 16: 618–621.
33. Rio E, Moseley L, Purdam C, et al. The pain of tendinopathy: physiological or pathophysiological? *Sports Med* 2014; 44: 9–23.