Reliability of Evaluating Achilles tendon blood flow assessed with Doppler Ultrasound Advanced Dynamic Flow

***Authors**

Lucie Risch

Monique Wochatz

Janin Messerschmidt

Tilman Engel

Frank Mayer

Michael Cassel

University Outpatient Clinic, Sports Medicine and Sports Orthopaedics,
University of Potsdam, Germany

***Corresponding Author.**

Lucie Risch

University Outpatient Clinic, Sports Medicine and Sports Orthopaedics,
University of Potsdam, Germany

Email: lrisch@uni-potsdam.de

Abstract

Reliability of quantifying intratendinous blood flow (IBF) assessed with the high-sensitive Doppler ultrasound Advanced Dynamic Flow (ADF) has not been examined, yet. Therefore, this study aimed to investigate intra- and inter-observer reliability of evaluating Achilles tendon IBF assessed with ADF using established scoring systems. Three investigators evaluated IBF in 67 recordings in a test-retest design applying the Ohberg score (OS), a modified Ohberg score (MOS), and a counting score (CS). Intra- and inter-observer agreement for OS and MOS was analysed with Cohen's Kappa and Fleiss Kappa (absolute), Kendall's tau b and Kendall's Coefficient of Concordance (Kendall's W) (relative). Reliability of CS was analysed with ICC 2.1 and 3.1, standard error of measurement (SEM) and Bland-Altman analysis (bias and limits of agreement (LoA)). Intra- and inter-observer agreement (absolute/relative) ranged from 0.61-0.87/0.87-0.95 and 0.11-0.66/0.76-0.89 for OS, and from 0.81-0.87/0.92-0.95 and 0.64-0.80/0.88-0.93 for MOS, respectively. CS revealed an intra-observer ICC of 0.94-0.97, SEM of 1.0-1.5, and bias of -1, with LoA of 3-4 vessels. Inter-observer ICC for CS ranged from 0.91-0.98, SEM of 1.0-1.9, and a bias 0, with LoA of 3-5 vessels. MOS and CS showed excellent reliability and seem convenient for research and clinical practice. OS revealed decent intra- but unexpected low inter-observer reliability and therefore cannot be recommended.

Keywords: Reliability, Intratendinous blood flow, Advanced Dynamic Flow

1. Introduction

Sonographically detectable intratendinous blood flow (IBF) is frequently associated with tendon pathology due to its' occurrence in 47-88% of Achilles tendinopathy patients [1]. However, a mere association to pathology is questionable [1,2], since IBF has been recently also found in up to 35% of asymptomatic Achilles tendons (AT) [3,4], and has shown an increased presence in response to exercise [5–7]. A differentiation between physiological (1-2 vessels) and pathological IBF (>2 vessels) has been suggested [2] but lacks consistent evidence [8]. Overall, it remains debatable if sonographically detectable IBF is an indicator of pathology [9,10], implies a predisposition to pathology [8], or can also be considered a physiological finding [2,5–7,11]. Further exploration of the role of IBF requires a precise, reliable and practicable assessment that accurately determines the extent of IBF. Hence, the scoring procedure is one essential task.

Various qualitative, semi-quantitative and quantitative scoring systems are available for evaluating the degree of IBF commonly assessed by power or colour Doppler ultrasound (PDU, CDU) [12]. Scores must account for a wide range of IBF since the amount can vary from occasionally present single, tiny vessels up to a great extent of complex vascularity throughout the tendon [13]. Originally, the first scoring system has been defined by Ohberg et al [14] for CDU imaging, grading IBF from 0 up to 4+ according to the presence of vessels inside the tendon. This score has been modified several times for PDU and CDU imaging, adapting categorisations by altering the localisation and amount of vessels or defining higher grades based on percentage area of the tendon covered with colour [6,8,11,15]. Another approach has been the qualitative grading of "no", "mild", "moderate", and "severe" IBF [16] comparable to scores used in rheumatology [17]. Others have evaluated IBF in CDU/PDU recordings by determining total vessel length (mm) [13], counting the total number of vessels [18], or by quantifying colour pixels in the Doppler ultrasound (DU) image [2,5,19,20]. Reliability studies have reported inter-observer ICC for PDU examinations with modified Ohberg score of 0.85 [15], and by estimating vessel length in CDU recordings of ICC 0.84-0.85 [13]. Qualitative scoring has revealed decent intra- (Spearman's r=0.75-0.85) and inter-observer (Spearman's r=0.76-0.92) reliability for CDU imaging [16]. However, visualization and discrimination of small vessels applying conventional Doppler modes are limited due to low frame rate, over painting of vessel walls (blooming effect) and presence of artefacts [21]. Large coloured areas in the ultrasound image might consist of several smaller vessels rather than just one [7]. In consequence, the exact quantification of IBF and especially a differentiation between low amount of vessels, e.g. discrimination between potential physiological and pathological IBF [2], is limited.

A high-sensitive, recently developed broadband colour Doppler "Advanced Dynamic Flow" (ADF) with US-imaging comparable to B-scan quality, higher resolution and frame rate, has shown enhanced precision in distinctly depicting and discriminating vessels compared to conventional DU [21,22]. Its' applicability in IBF examinations has been recently shown [4,8,23]. However, taking advantage of this improved imaging quality requires an adequate and reliable quantitative scoring system [23] which has not been investigated, yet. Therefore, the aim of this study was to investigate the intra- and inter-observer reliability of evaluating DU recordings of AT IBF examined with ADF by applying three established scoring systems used for conventional DU modes: the original Ohberg Score (OS) [14], a modified Ohberg score described by Hirschmüller et al. [3] (MOS), and counting the number of intratendinous vessels [18] (CS). It is hypothesized that all 3 procedures reveal good intra-and inter-observer agreement.

2. Methods

2.1 Participants

IBF was assessed in several areas with blood flow throughout the whole left and right AT (from tendon insertion to musculotendinous junction) of fourteen subjects (7 males, 7 females; 39±12 years; 1.76±0.10 m; 74±15 kg) with acute or chronic Achilles tendinopathy (VISA-A Score 72±15 [24]), resulting in 67 recordings. The participants were recruited from the University outpatient clinic. All participants gave their written informed consent. The study was approved by the local ethics committee.

2.2 Study design

Sixty-seven ultrasound recordings of AT IBF were evaluated by three blinded investigators (I1, I2, I3). For an equal understanding of the three scoring systems, all investigators were instructed together prior to the study. To investigate inter-observer reliability, each investigator scored the same randomized recordings three times applying the different scoring systems in randomized order (M1). To investigate intra-observer reliability, the procedure was repeated (M2) with a different randomized order of scoring systems and recordings (Figure 1). Measurements were separated by intervals of at least two days.

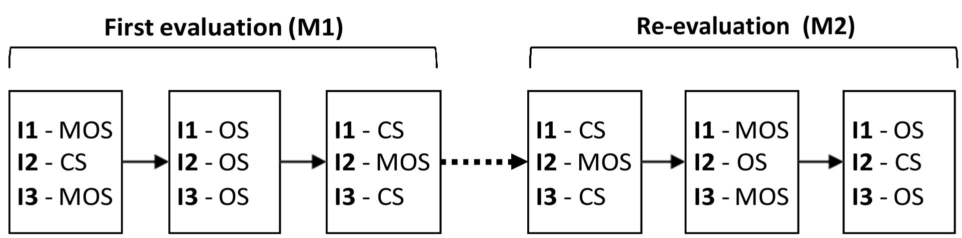


Figure 1. Flowchart of study protocol, I1-I3= investigators 1, 2, and 3; OS= Ohberg Score, MOS= modified Ohberg Score, CS= Counting Score

2.3 Ultrasound recordings

IBF images were recorded with a high-resolution ultrasound device (Xario SSA-660A, Toshiba) using a multi-frequency linear transducer at 14 MHz (PLT-1204AT). The Doppler mode ADF had standardized pre-settings: color gain 40-42, color velocity 1.5 cm/s, and pulse repetition frequency 12.5 kHz. The box size of the region of interest (ROI) was 2.0 cm wide and 1.5 cm deep. The analyzed ADF recordings consisted of video sequences (5 s) saved as AVI files (Figure 2).

2.4 Data analysis

The investigators applied three scoring systems for quantifying IBF in the ultrasound recordings: (1) the OS grading IBF from 0 up to 4+ according to the appearance of vessels inside a tendon. Grade 0 referred to no intratendinous vessel. One or two small vessels located in the frontal part of the tendon were scored 1+. Several irregular vessels throughout the tendon were quantified as 2+ to 4+ [14]; (2) a MOS presented by Hirsch Müller et al. [8] defining six different grades as follows: 0= no vessels visible, 1 = 1–2 vessels within the region of interest (ROI), 2 = 3–5 vessels within the ROI, 3= vessels in up to 30% of the ROI, 4 = vessels in 30%–50% of the ROI, 5= vessels in >50% of the ROI; (3) CS, counting the total number of vessels, every branch representing a single vessel.

2.5 Statistical Analysis

Statistical analysis was performed using Microsoft Excel 2010 and IBM SPSS Statistics Version 20. Significance level was set to α= 0.05. Descriptive statistics are presented as median and range, and mean ± standard deviation (SD). Intra- and pairwise inter-observer reliability of OS and MOS were analyzed by calculating Cohen's Kappa coefficient for absolute agreement and Kendall's tau b correlation coefficient for relative agreement, taking the degree of deviation into account (main result) [25]. Inter-observer comparison between all three investigators was analyzed with Fleiss Kappa coefficient (absolute agreement) and Kendall's coefficient of concordance (Kendall's W, relative agreement). Intra- and pairwise inter-observer reliability for CS was analyzed with the ICC 2.1 and 95% confidence interval (CI), standard error of measurement (SEM, square root of the mean square error term in repeated measures ANOVA [26]) and by performing a Bland-Altman analysis with bias and limits of agreement (LOA, bias±1.96*SD). Inter-observer comparison between all three investigators was analyzed with ICC 3.1 and 95% CI, and SEM. Coefficients are interpreted according to Landis and Koch [27] as "poor" (<0.0), "slight" (0.0-0.20), "fair" (0.21-0.40), "moderate" (0.41-0.60), "substantial"

(0.61-0.80), and “excellent” (0.81-1.00).

3. Results

The overall median score in M1 and M2 was 3 and 4 (range 0-4) for OS, 2 and 2 (range 0-5) for MOS respectively. Overall mean CS was 6±6 (range 0-34) in M1 and 7±7 (range 0-34) in M2. Distributions of IBF scores (M1) for each grading system are presented in Figures 3a, b, c.

The OS resulted in substantial to excellent absolute (0.61-0.87) and excellent relative (0.87-0.95) intra-observer agreement, slight to substantial absolute (0.11-0.66), and substantial to excellent relative (0.76-0.92) inter-observer agreement (Table 1 and 2). The MOS showed overall higher consistency than OS with excellent absolute (0.81-0.87) and relative (0.92-0.95) intra-observer agreement, and substantial absolute (0.64-0.80) and excellent relative (0.88-0.97) inter-observer agreement (Table 1 and 2). CS revealed highest correlations with excellent intra-observer agreement (ICC 0.94-0.97), SEM of 1.0-1.5, a systematic error of -1 and a random error of 3-4 vessels. Also, inter-observer reliability was excellent (ICC 0.91-0.98), SEM of 1.0-1.9, with no systematic error between observers and a random error that ranged between 3-5 vessels (Table 1 and 2, Figure 4 a, b)

Table 1. Intra- and pairwise inter-observer reliability for OS, MOS, and C

		OS		MOS		CS		
		Cohens Kappa	Kendall’s tau b	Cohens Kappa	Kendall’s tau b	ICC (95% CI)	SEM	Bias ± LoA
Intra-rater	I 1	0.61	0.87	0.87	0.95	0.94 (0.90 -0.96)	1.47	-1±4
	I 2	0.87	0.95	0.87	0.95	0.97 (0.94-0.99)	0.99	-1±3
	I 3	0.77	0.90	0.81	0.92	0.97 (0.94-0.99)	1.29	-1±3
Inter-rater M1	I1 vs. I2	0.30	0.82	0.68	0.91	0.94 (0.91-0.96)	1.50	0±4
	I1 vs. I3	0.11	0.77	0.64	0.89	0.91 (0.86-0.94)	1.85	0±5
	I2 vs. I3	0.66	0.89	0.78	0.92	0.94 (0.91-0.96)	1.61	0±4
Inter-rater M2	I1 vs. I2	0.61	0.86	0.78	0.91	0.94 (0.90-0.96)	1.60	0±4
	I1 vs. I3	0.29	0.76	0.64	0.88	0.95 (0.93-0.97)	1.45	0±4
	I2 vs. I3	0.54	0.83	0.80	0.93	0.98 (0.96-0.99)	1.04	0±3

OS= Ohberg Score, MOS= Modified Ohberg Score, CS= Counting Score; I1, I2, I3= investigators 1, 2, 3

Table 2. Inter-observer reliability of 3 observers for OS, MOS, and CS

	OS		MOS		CS	
	Fleiss Kappa	Kendall's W	Fleiss Kappa	Kendall's W	ICC (95% CI)	SEM
Inter-observer M1	0.34	0.92	0.70	0.97	0.93 (0.90-0.96)	1.66
Inter-observer M2	0.46	0.92	0.74	0.97	0.96 (0.94-0.99)	1.39

OS = Ohberg Score, MOS = Modified Ohberg Score, CS = Counting Score

Figure 2. Exemplary Doppler-ultrasound still frame from video sequence (a) scored OS: 4, MOS: 5, CS: 25-34 (b) scored OS: 2-4, MOS: 3, CS: 5-10

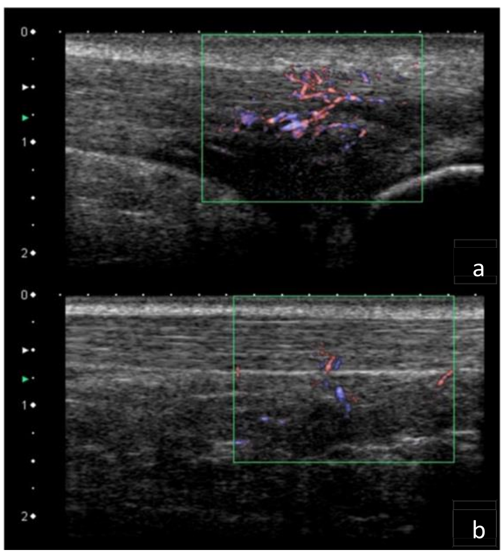
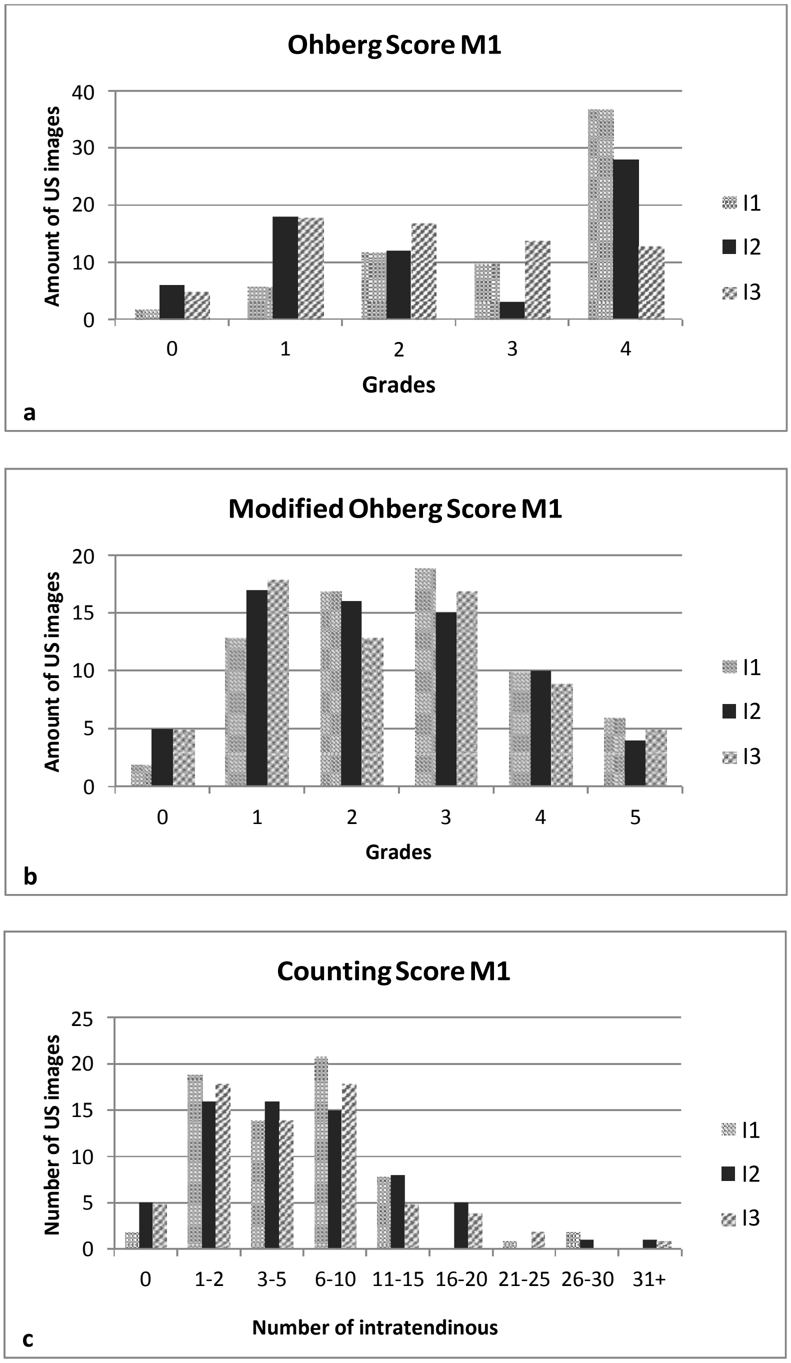
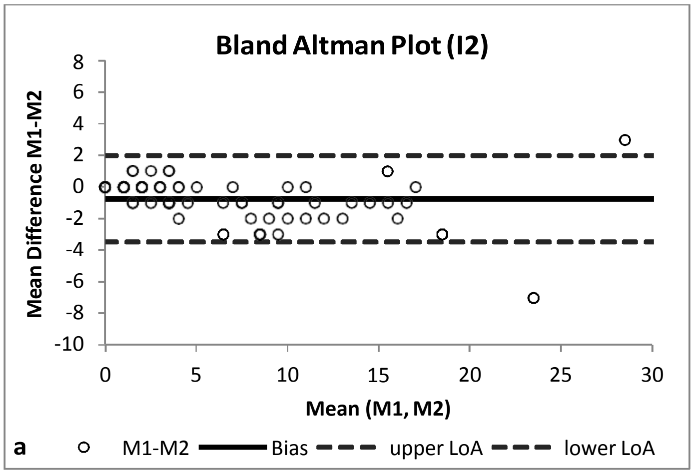
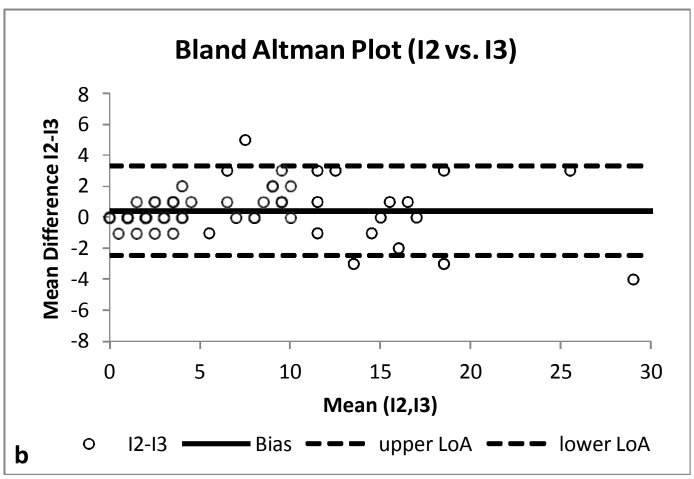


Figure 3. Distribution of Scores for (a) OS, (b) MOS and (c) CS. CS results are grouped for better display



Figures 4. Exemplary Bland Altman plots for (a) intra-observer and (b) pairwise inter-observer reliability of CS





4. Discussion

The aim of this study was to investigate reliability of three scoring procedures for evaluating IBF in ATs assessed with DU ADF. The CS showed highest agreement for both intra- and inter-observer reliability, followed by very good results for MOS with slightly higher intra- than inter-observer reliability. OS revealed decent intra- but unexpected low inter-observer agreement. Both CS and MOS seem convenient for evaluating ADF examinations in research and clinical practice and showed even higher reliability than previously reported for scoring systems from conventional Doppler modes [13,15]. The original OS on the other hand cannot be recommended for ADF imaging.

The three scoring procedures applied for ADF imaging were chosen due to their assumed practicability for clinical assessment. The CS seems advantageous for quantifying the absolute amount of IBF throughout the tendon during static and dynamic recordings. According to our results, it can be considered very reliable, with a measurement error of 1-2 vessels, and observer-independent, since there was no relevant difference in agreement between intra- and inter-observer scoring. Due to the excellent reliability it seems sensitive to slight changes in number of vessels even when scoring complex vascularity, however, remains ignorant of their magnitude in terms of vessel size and length. Nevertheless, this scoring procedure allows for easy, immediate and absolute IBF quantification and therefore seems adequate for application in clinical practice.

The MOS seems feasible to grade the amount of vessels in a specific region of interest of an image, with the advantage of discriminating up to a number of 5 single vessels, and differentiating higher amount of blood flow in percentage area of the tendon by also regarding the magnitude of vascularisation [8]. A disadvantage is the lack of precision to determine smaller changes in higher scores and its impracticability to evaluate blood flow in the whole tendon. The MOS showed slightly higher correlations for intra-observer than inter-observer reliability. Therefore, it is recommended to perform evaluation using the MOS by the same investigator. Obviously, the relative agreement showed higher results than the absolute agreement, indicating that deviation in the non-matching scores was very low.

The OS was originally designed to assess a lower number of vessels throughout the tendon in CDU images [14] since Doppler modes in the past have been much less sensitive and precise in visualizing discrete vessels compared to improved Doppler modes such as ADF [2]. Nevertheless, the assessment of OS grades is based on categorizing number of vessels and therefore was expected to also reveal good reliability. However, the OS showed lowest correlations with better intra- than inter-observer reliability. The high relative agreement showed that disagreeing scores did not deviate much. However, absolute inter-observer agreement results were so low that the OS cannot be considered applicable for practice. It is speculated that the low inter-observer reliability indicates a lack of accurate, consistent definition of OS grades and might be one reason for the numerous adaptations this score has experienced in the past [5, 9, 10, and 14].

Currently available scores are usually based on number of vessels, vessel length or determining area of the tendon covered with colour using special software (pixel count, colour fraction measurement) [2,8,13,18–20]. However, the clinical relevance of IBF presence and its' magnitude (vessel length and size) remains unclear and needs further investigation [1,13,28]. ADF presents enhanced imaging quality with less blooming effect, superior "discrimination from neighboring vessels" and more precision in "following course of the vessels" compared to conventional Doppler modes [21]. This might enable a more exact quantification and monitoring of IBF in research and clinical practice and might also help to define a proposed threshold to differentiate between physiological and pathological blood flow [1,2,5]. Since counting the number of vessels is irrespective of determining their extent, we propose to combine the CS with quantitative measures, e.g. colour fraction measurement [20], to explore the role of IBF in the context of tendon pathology or physiology in further studies.

5. Conclusion

CS and MOS both showed high reliability and seem convenient for use in research and clinical practice, while OS cannot be recommended in its original version. CS revealed highest reliability independent of observer, and seemed the most precise score for evaluating amount of vessels in ADF recordings. Since the score lacks to display the magnitude of vessels, we propose to combine quantitative evaluation in terms of number and size of vessels in future studies to explore the role of IBF in the context of tendon pathology and physiology. These features might enable a more accurate IBF assessment, a more precise monitoring of changes over time, and the determination of a suggested threshold between pathological and physiological blood flow.

The author declares no conflict of interest regarding publication of this paper.

References

1 Tol JL, Spiezia F, Maffulli N. Neovascularization in Achilles tendinopathy: have we been chasing a red herring? Knee Surg Sports Traumatol Arthrosc 2012; 20: 1891–1894

2 Boesen AP, Boesen MI, Torp- Pedersen S, Christensen R, Boesen L, Hölmich P, Nielsen MB, Koenig MJ, Hartkopp A, Ellegaard K, Bliddal H, Langberg H. Associations between abnormal ultrasound color Doppler measures and tendon pain symptoms in badminton players during a season: a prospective cohort study. Am J Sports Med 2012; 40: 548–555

3 Hirschmüller A, Frey V, Deibert P, Konstantinidis L, Mayer F, Südkamp N, Helwig P. [Achilles tendon power Doppler sonography in 953 long distance runners - a cross sectional study]. Ultraschall Med 2010; 31: 387–393

4 Cassel M, Baur H, Hirschmüller A, Carlsohn A, Fröhlich K, Mayer F. Prevalence of Achilles and patellar tendinopathy and their association to intratendinous changes in adolescent athletes. Scand J Med Sci Sports 2014; 1–9

5 Boesen MI, Koenig MJ, Torp- Pedersen S, Bliddal H, Langberg H. Tendinopathy and Doppler activity: the vascular response of the Achilles tendon to exercise. Scand J Med Sci Sports 2006; 16: 463–469

6 Boesen MI, Boesen A, Koenig MJ, Bliddal H, Torp-Pedersen S. Ultrasonographic investigation of the Achilles tendon in elite badminton players using color Doppler. Am J Sports Med 2006; 34: 2013–2021

7 Malliaras P, Richards PJ, Garau G, Maffulli N. Achilles tendon Doppler flow may be associated with mechanical loading among active athletes. Am J Sports Med 2008; 36: 2210–2215

8 Hirschmüller A, Frey V, Konstantinidis L, Baur H, Dickhuth H-H, Südkamp NP, Helwig P. Prognostic value of Achilles tendon Doppler sonography in asymptomatic runners. Med Sci Sports Exerc 2012; 44: 199–205

9 Ohberg L, Lorentzon R, Alfredson H. Neovascularisation in Achilles tendons with painful tendinosis but not in normal tendons: an ultrasonographic investigation. Knee Surg Sports Traumatol Arthrosc 2001; 9: 233–238

10 Alfredson H Ohberg L, Forsgren S. Is vasculo-neural ingrowth the cause of pain in chronic Achilles tendinosis? An investigation using ultrasonography and colour Doppler, immunohistochemistry, and diagnostic injections. Knee Surg Sports Traumatol Arthrosc 2003; 11: 334–338

11 de Vos R-J, Weir A, Cobben LPJ, Tol JL. The value of power Doppler ultrasonography in Achilles tendinopathy: a prospective study. Am J Sports Med 2007; 35: 1696– 1701

12 Yang X, Pugh ND, Coleman DP, Nokes LDM. Are Doppler studies a useful method of assessing neovascularization in human Achilles tendinopathy? A systematic review and suggestions for optimizing machine settings. J

Med Eng Technol 2010; 34: 365–372

13 Cook JL, Ptazsnik R, Kiss ZS, Malliaras P, Morris ME, De Luca J. High reproducibility of patellar tendon vascularity assessed by colour Doppler ultrasonography: a reliable measurement tool for quantifying tendon pathology. Br J Sports Med 2005; 39: 700–703

14 Ohberg L, Alfredson H. Ultrasound guided sclerosis of neovessels in painful chronic Achilles tendinosis: pilot study of a new treatment. Br J Sports Med 2002; 36: 173–175; discussion 176–177

15 Sengkerij PM, de Vos R-J, Weir A, van Weelde BJG, Tol JL. Interobserver reliability of neovascularization score using power Doppler ultrasonography in midportion achilles tendinopathy. Am J Sports Med 2009; 37: 1627–1631

16 Sunding K, Fahlström M, Werner S, Forssblad M, Willberg L. Evaluation of Achilles and patellar tendinopathy with greyscale ultrasound and colour Doppler: using a four-grade scale. Knee Surgery, Sport Traumatol Arthrosc 2014;

17 Weidekamm C, Köller M, *Weber* M, Kainberger F. Diagnostic value of high-resolution B-mode and Doppler sonography for imaging of hand and finger joints in rheumatoid arthritis. Arthritis Rheum 2003; 48: 325–333

18 Richards PJ, Win T, Jones PW. The distribution of microvascular response in Achilles tendonopathy assessed by colour and power Doppler. Skeletal Radiol 2005; 34: 336–342

19 Malliaras P, Chan O, Simran G, Martinez de Albornoz P, Morrissey D, Maffulli N. Doppler ultrasound signal in Achilles tendinopathy reduces immediately after activity. Int J Sports Med 2012; 33: 480–484

20 Beyer R, Kongsgaard M, Hougs Kjaer B, Ohlenschlaeger T, Kjaer M, Magnusson SP. Heavy Slow Resistance Versus Eccentric Training as Treatment for Achilles Tendinopathy: A Randomized Controlled Trial. Am J Sports Med 2015; 43: 1704–1711

21 Heling K S, Chaoui R, Bollmann R. Advanced dynamic flow -- a new method of vascular imaging in prenatal medicine. A pilot study of its applicability. Ultraschall Med 2004; 25: 280–284

22 Fittschen M, Reinhard I, Wellek S, Friedrichs S, Bahlmann F. Advanced dynamic Doppler flow of the pulmonary artery in a normal population: reference values from 18 to 41 weeks of gestation calculated by automatic Doppler waveform analysis. Arch Gynecol Obstet 2014; 289: 973–980

23 Risch L, Cassel M, Messerschmidt J, Intziegianni K, Fröhlich K, Kopinski S, Mayer F. Is Sonographic Assessment of Intratendinous Blood Flow in Achilles Tendinopathy Patients Reliable? Ultrasound Int Open 2016; epub ahead of print.

24 Robinson JM, Cook JL, Purdam C,Visentini PJ, Ross J, Maffulli N,Taunton JE, Khan KM. The VISA- A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. Br J Sports Med 2001; 35: 335–341

25 Bortz J, Lienert GA. Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben. 3rd ed. Heidelberg: Springer, 2008

26 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998; 26: 217–238

27 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174

28 Cook JL, Malliaras P, De Luca J, Ptasznik R, Morris M. Vascularity and pain in the patellar tendon of adult jumping athletes: a 5 month longitudinal study. Br J Sports Med 2005; 39: 458–461; discussion 458– 461