Osteoarthritis and Cartilage 25 (2017) 790–798

Osteoarthritis and Cartilage

OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Multimodality scoring of chondral injuries in the equine fetlock joint *ex vivo*

J.K. Sarin † ‡ *, H. Brommer §, D. Argüelles ||, P.H. Puhakka † ‡, S.I. Inkinen †, I.O. Afara † ‡ ¶, S. Saarakkala # †† ‡‡, J. Töyräs † ‡

† Department of Applied Physics, University of Eastern Finland, Kuopio, Finland
‡ Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland
§ Department of Equine Sciences, Utrecht University, Utrecht, The Netherlands
|| Veterinary Teaching Hospital, School of Veterinary Medicine, University of Helsinki, Helsinki, Finland
¶ Department of Electrical and Computer Engineering, Elizade University, Ondo, Nigeria
# Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
†† Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
‡‡ Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland

ARTICLE INFO

Article history:
Received 16 May 2016
Accepted 1 December 2016

Keywords:
Cartilage
Reproducibility
Reliability
Injury
Scoring

SUMMARY

Objective: We investigate the potential of a prototype multimodality arthroscope, combining ultrasound, optical coherence tomography (OCT) and arthroscopic indentation device, for assessing cartilage lesions, and compare the reliability of this approach with conventional arthroscopic scoring *ex vivo*.

Design: Areas of interest (AIs, $N = 43$) were selected from equine fetlock joints ($N = 5$). Blind-coded AIs were independently scored by two equine surgeons employing International Cartilage Repair Society (ICRS) scoring system via conventional arthroscope and multimodality arthroscope, in which high-frequency ultrasound and OCT catheters were attached to an arthroscopic indentation device. In addition, cartilage stiffness was measured with the indentation device, and lesions in OCT images scored using custom-made automated software. Measurements and scorings were performed twice in two separate rounds. Finally, the scores were compared to histological ICRS scores.

Results: OCT and arthroscopic examinations showed the highest average agreements (55.2%) between the scoring by surgeons and histology scores, whereas ultrasound had the lowest (50.6%). Average intraobserver agreements of surgeons and interobserver agreements between rounds were, respectively, for conventional arthroscope (68.6%, 69.8%), ultrasound (68.6%, 68.6%), OCT (65.1%, 61.7%) and automated software (65.1%, 59.3%).

Conclusions: OCT imaging supplemented with the automated software provided the most reliable lesion scoring. However, limited penetration depth of light limits the clinical potential of OCT in assessing human cartilage thickness; thus, the combination of OCT and ultrasound could be optimal for reliable diagnostics. Present findings suggest imaging and quantitatively analyzing the entire articular surface to eliminate surgeon-related variation in the selection of the most severe lesion to be scored.

Introduction

Evaluation of articular cartilage injuries by traditional imaging methods, such as radiography and magnetic resonance imaging, has been shown to correlate poorly with an arthroscopic examination[1–4]. Hence, decision on the optimal treatment is often based on arthroscopic findings during the surgery. Several scoring systems have been proposed for the evaluation of cartilage lesions in arthroscopy, including International Cartilage Repair Society –

* Address correspondence and reprint requests to: J.K. Sarin, Department of Applied Physics, University of Eastern Finland, POB 1627, FI-70211, Kuopio, Finland.
E-mail addresses: jaakko.sarin@uef.fi (J.K. Sarin), H.Brommer@uu.nl (H. Brommer), darguelles@uco.es (D. Argüelles), Pia.Puhakka@essote.fi (P.H. Puhakka), satu.inkinen@uef.fi (S.I. Inkinen), isaac.afara@uef.fi (I.O. Afara), simo.saarakkala@oulu.fi (S. Saarakkala), juha.toyras@uef.fi (J. Töyräs).

http://dx.doi.org/10.1016/j.joca.2016.12.007
1063-4584/

ICRS[5], Outerbridge[6], and French arthroscopic society – SFA[7]. However, the validity of these classifications is restricted by surgeons' subjectivity in determining the depth of cartilage lesion and chondral softening, resulting in poor intraobserver and interobserver reproducibility[8–10]. Therefore, more quantitative and objective approaches, such as acoustic and optical techniques, as well as computer-assisted cartilage lesion scoring, are desirable[9,11–15].

Several non-destructive techniques, including intra-articular ultrasound (US)[13,16], optical coherence tomography (OCT)[14,17], near infrared (NIR) spectroscopy[15,18,19] and arthroscopic indentation[20,21], have been proposed for arthroscope guided quantitative and objective evaluation of cartilage integrity. US and OCT provide cross-sectional images of tissue structure, but are also complementary to each other: the superior resolution of OCT enables the high-resolution characterization of the articular surface, whereas US provides detailed information on the inner structures of cartilage[22] and enables subchondral bone evaluation[13]. Clinical intravascular US and OCT catheters are suitable for imaging narrow joint cavities and have been shown to be feasible for enhancing the accuracy of cartilage lesion scoring during arthroscopy[13,14,23]. However, the reproducibility of lesion scoring based on the US and OCT images is also restricted by the subjectivity of a surgeon[23].

Visual arthroscopic evaluation is supported by assessment of cartilage stiffness by palpating the articular surface with an arthroscopic probe, thus enabling the detection of chondral softening and cartilage flaps. A number of hand-held devices have been developed for determining cartilage stiffness during the arthroscopic surgery[20,24]. The stiffness of cartilage depends on its biochemical composition; thus, subtle compositional changes can alter the biomechanical response of cartilage[25,26]. Furthermore, cartilage stiffness is site-dependent[27–29] and may not be directly evaluated with imaging methods such as US and OCT.

In this study, we applied a prototype multimodality arthroscopic system, combining US, OCT and arthroscopic indentation device[20], for quantitative assessment of cartilage injuries. We investigated the intraobserver and interobserver reproducibility of the imaging techniques for assessing different levels of cartilage injury, and compared the outcome with conventional arthroscopic evaluation based on the ICRS scoring. Furthermore, OCT images obtained by the surgeons were automatically scored by custom-made software. Biomechanical assessment with arthroscopic device was compared with that using a laboratory indentation system. We hypothesized that the scoring of cartilage lesions from US and OCT images provides better reproducibility than that via conventional arthroscopy. Additionally, software-based scoring was expected to offer superior reproducibility compared to the individual techniques.

Materials and methods

Sample preparation

Fetlock joints were extracted from mature equine cadavers ($N = 5$) and stored at −20°C until required for the experiment (Fig. 1). The joints were opened prior to the experiment to guarantee inclusion of cartilage defects with various severity and to permit the marking of the areas of interest (AIs), thus ensuring well-confined AIs and reliable location tracking for the blind-coded scoring. Sample preparation was performed by researchers to eliminate any surgeon-related bias. Out of the two experienced board-certified equine surgeons (~900 arthroscopies, Diplomate European College of Veterinary Surgeons), the more experienced surgeon (~500 arthroscopies) selected the AIs ($N = 44$, Area ≈ 0.8 cm^2) based on visual evaluation to include both intact and damaged cartilage regions. Distinct lesions were centralized within the AIs. Throughout all the measurements, samples were submerged in room temperature phosphate-buffered saline (PBS).

The most severe lesion within each blind-coded AI was independently scored twice by both surgeons according to the ICRS cartilage injury classification system[5]: normal cartilage is classified as grade 0; fibrillated or superficially lacerated cartilage as grade 1; deeper lacerations but defects extending less than 50% of cartilage thickness as grade 2; defects extending deeper than 50% of cartilage thickness as grade 3 and defects extending into subchondral bone as grade 4. The scoring system was reviewed in detail with both surgeons prior to the experiments; furthermore, the scoring guidelines were visible for the surgeon throughout the measurements. During the lesion evaluation, the measurement instruments were submerged in PBS.

Conventional and prototype multimodality arthroscopes

First, cartilage was evaluated by examining the whole AI with an arthroscopic system, including a conventional clinical arthroscope (lens inclination = 30°, diameter = 4 mm, 28731 BWA, Karl Storz, GmbH & Co, Germany), a light source (Xenon XL, Smith & Nephew, Dyonics, Memphis, Tennessee, USA) and an arthroscopic hook probe (28145S (90°), Karl Storz). As a result of removing the joint capsule and other overlying tissues, the surgeons had unrestricted access to the defects from an optimal angle and distance. Subsequently, a prototype multimodality arthroscope [Fig. 2(A), (D)], including clinical US (ClearView Ultra, Boston Scientific Corporation, Marlborough, MA, USA) and OCT systems ($\lambda = 1300 \pm 55$ nm, Ilumien PCI Optimization System, St. Jude Medical, St. Paul, MN, USA), was utilized in similar examination. Intravascular catheters (US: Atlantis SR Pro, $f_c = 40$ MHz, Boston Scientific Corporation; OCT: C7 Dragonfly, St. Jude Medical) were used for real-time cross-sectional visualization and scoring of cartilage integrity. The quality of the imaging live feed was monitored throughout the experiment to ensure optimal image acquisition for scoring with the automatic software. The superior resolution of OCT (axial resolution ≤20 μm, lateral resolution 25–60 μm) enables high fidelity imaging of cartilage surface [Fig. 2(B), (E)], while US (axial resolution ≥43 μm) is capable of penetrating deep into the tissue, allowing better assessment of the cartilage–bone interface [Fig. 2(C), (F)]. Both surgeons were relatively unfamiliar with the novel imaging modalities (US and OCT); however, the modalities are easily adaptable and require no extensive practicing.

In the multimodality arthroscope, the imaging catheters were attached on the opposite side than the indenter of the Artscan 200 indentation device; thus, the prototype was rotated 180° to change between the imaging and indentation modalities. Throughout imaging the distance between the catheter and cartilage surface was kept minimal to ensure optimal image quality as the lateral resolution decreases with increasing distance from the catheters. The shape (curvature) of articular surface had no significant effect on the imaging modalities as no quantitative reflection or backscattering parameters were determined. Furthermore, the geometry of the arthroscopic indentation device limits its application at extremely concave surfaces; however, all the AIs could be evaluated.

Biomechanical testing

The stiffness of cartilage within AIs was determined with a hand-held arthroscopic indentation device – Artscan 200 (Artscan Oy, Helsinki, Finland)[20], which was part of the prototype multimodality arthroscope. The functionality of the device has been described previously[21], and is summarized here. The reference plate is gently pressed against the cartilage surface with minimum threshold force of 5 N, allowing a spherical indenter protruding

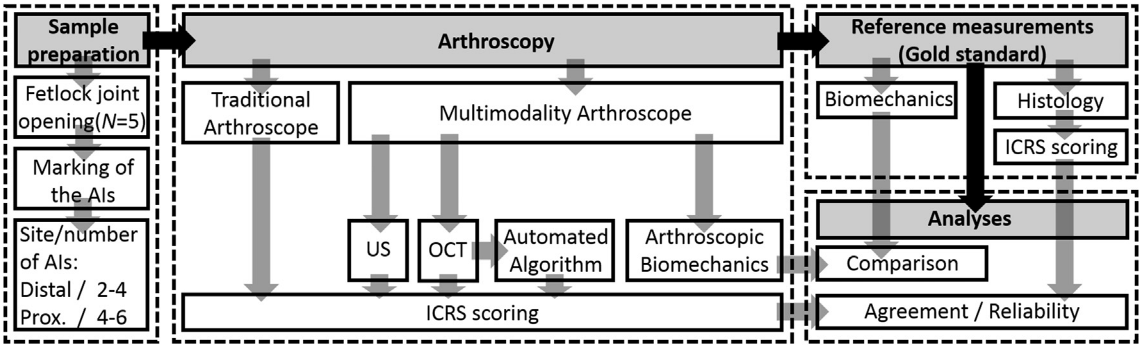

Fig. 1. Study design. The arthroscopic ICRS scoring was performed by two experienced equine surgeons two times, and the histological scoring by three evaluators three times.

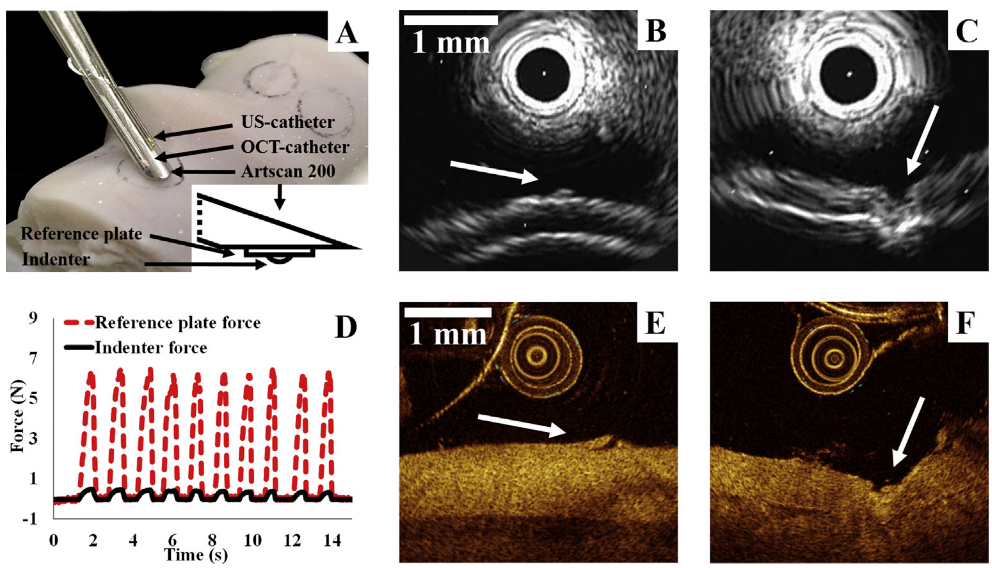

Fig. 2. The alignment (A) and output signal (D) of the hand-held indentation device for probing the mechanical properties of articular cartilage are shown. Higher resolution of OCT enables enhanced visualization of cartilage surface (B vs E), whereas US enables superior detection of subchondral bone (C vs F). Same lesions are indicated with arrows.

from the reference plate to indent the cartilage and measure the force by which the tissue resists the deformation [Fig. 2(A), (D)]. Ten successive indentations were performed on the cartilage surface; the measurement was repeated three times for each AI per scoring round by both surgeons.

The mechanical competence of cartilage was assessed by determining the instantaneous modulus ($E_{Artscan}$) based on the average of five maximum indenter forces. The modulus was determined based on the Hayes' elastic model of indentation[30]:

$$E = \frac{F(1-\nu)RX}{4a^3\kappa},$$

where F is the measured indenter force, ν is the Poisson's ratio ($\nu = 0.5$), R is the indenter radius of curvature, α is the height of the indenter, and X and κ are theoretical correction factors[30]. The spherical indenter height and radius of curvature were 0.13 mm and 0.35 mm, respectively. Out of the three repetitions the lowest modulus value was excluded to rule out the measurements with possibly imperfect contact between the indenter and cartilage.

Additionally, instantaneous modulus (E_{Lab}) was determined using a laboratory material testing system, comprising of a load cell with a force resolution of 5 mN (Sensotec, Columbus, OH, USA) and an actuator with a displacement resolution of 0.1 μm (PM500-1 A, Newport, Irvine, CA, USA). A plane-ended cylindrical indenter ($d = 0.53$ mm) was aligned perpendicular to the sample surface, and driven into contact with the surface with a pre-stress of 12.5 kPa. A 7.5% strain step, relative to the cartilage thickness, was performed at a strain rate of 100%/s. E_{Lab} was calculated along Hayes *et al.* by assuming $\nu = 0.5$.

Automatic software for lesion scoring

The lesion scoring software utilized in this study has been previously introduced by te Moller *et al.* 2016[31]. In summary, the software determines the tidemark of calcified and non-calcified cartilage and cartilage surface from an OCT image in order to calculate the non-calcified cartilage thickness. The roughness of cartilage surface (OCT roughness index, ORI[32]) is calculated to distinguish between ICRS scores 0 and 1, with a threshold of ORI = 8 μm. The maximum depth of cartilage lesion is determined and then compared to the tissue thickness to differentiate between ICRS scores 1 to 4. In an earlier study[31], a threshold of 8% loss of cartilage was applied for differentiating between ICRS scores 1 and 2. These thresholds were determined with an independent training set ($N = 148$)[23,31]. Other thresholds are based on the ICRS scoring system guidelines[5]. Unlike the previous semi-automated software presented by te Moller *et al.*, the current algorithm was modified to be fully automatic by selecting a 2 mm wide area perpendicularly under the OCT catheter as region of interest.

Histology and histological scoring

After the laboratory biomechanical measurements, the AIs were prepared for histological analyses. The samples were fixed in formalin, decalcified in EDTA and embedded in paraffin blocks. For each AI, three adjacent 5 μm sections were cut from three locations: along the central line of the AI, and two locations approximately 5 mm on both sides of the central line, and stained with Safranin-O. A blind-coded stained section of each AI containing the most severe histological changes was scored according the ICRS cartilage injury classification system by three independent evaluators in three blinded repetitions to determine the average gold standard score for each AI. A single AI was determined as an outlier and excluded due to poor location matching between the histological and OCT images.

Statistical analyses

All statistical analyses, including calculations of intraobserver, interobserver and inter-method agreements of scorings, were performed using IBM SPSS statistics software (Version 21, SPSS Inc., Chicago, USA). These parameters were selected based on the nature of the study; therefore, no correlations were investigated. For statistical analyses the histological scoring and biomechanical measurements using a laboratory material testing system were considered as the gold standards. Intraobserver and interobserver reproducibilities were assessed with the kappa (κ) coefficient, a chance-corrected estimate of agreement. Intraobserver reproducibility evaluates the internal reproducibility of a rater and interobserver reproducibility evaluates the reproducibility between raters. κ values 1.00–0.81 indicate excellent agreement, 0.80–0.61 substantial agreement, 0.60–0.41 moderate agreement, 0.40–0.21 fair agreement and 0.20–0.00 slight agreement[33]. Similarly, the inter-method reliability between surgeons' scores and histological scores (gold standard) was assessed. For evaluators of histological images, the intraclass correlation coefficient (ICC) was determined with two-way mixed model (absolute agreement) with median scores of each evaluator, with ICC(3,1) estimating the reliability of any one evaluators' scores and ICC(3,3) by using the mean score of all evaluators. Statistical significance was considered when $P < 0.05$. The 95% confidence intervals (CIs) were also determined. Standard deviation of the mean (SD) and coefficient of variation (CV_{RMS}) were also determined for Artscan 200 indentation measurements between repetitions and scoring rounds.

Results

The arthroscopic and histologic scores ranged from ICRS 0 to ICRS 3; thus, no lesions penetrating to subchondral bone were present. The scores ($N = 43$) assigned by both surgeons were systematically higher in the second scoring round (Table I), and the software scored the exact same OCT images systematically higher compared to the surgeons. Software scores were closest to the histology-based scores.

Inter-method reliability

OCT and arthroscopic examination were the most reliable techniques with the highest average agreements (55.2%) between scoring by the surgeons and histology-based scores (gold standard). Furthermore, the average agreement for the software was 53.5% and 50.6% for US. Overall, the inter-method agreement varied between 44.2% and 67.4% ($\kappa = 0.211$ to 0.540, Table II). In 38.6% of the cases, the histology-based scores by the independent evaluators were higher than the scores assigned by the surgeons based on the arthroscopic imaging techniques. Of these cases, 48.2% of the scores were ICRS 1. The difference between histology and imaging-based scores was greater than one in only 5.0% of cases. The ICC(3,1) and ICC(3,3) with 95% CIs for evaluators of histological images were 0.893 (0.827, 0.938) and 0.962 (0.935, 0.978), respectively ($P < 0.001$).

Intraobserver and interobserver reproducibilities

Intraobserver agreement of the imaging techniques ranged from 53.5 to 79.1% ($\kappa = 0.332$–0.660) (Table III). Overall, conventional arthroscope and US had the highest average intraobserver agreements (both 68.6%), whereas OCT and automatic software both had the average agreement of 65.1%. Surgeon 1 was more biased by the score of the previous technique than surgeon 2, with scores between the imaging methods totally agreeing in 81.4% and 57.0% of cases, respectively.

Surgeon 1 systematically assigned lower scores in both rounds when compared to surgeon 2, regardless of the imaging technique (Table I). Interobserver agreement of the imaging techniques ranged from 46.5 to 76.7% ($\kappa = 0.237$–0.615) (Table III) with the conventional arthroscope having the highest average interobserver agreement (69.8%) in comparison to US (68.8%), OCT (61.7%) and automatic software (59.3%). In OCT-based scoring, automatic software matched surgeons' scores in 63.4% and scored higher in 24.4% of cases.

Biomechanics

Average instantaneous modulus ($E_{Artscan}$) and its SD decreased between the two rounds with both surgeons ($P < 0.001$) (Table I). Significant correlation (63.6% and 63.3% for surgeons 1 and 2, respectively) was observed between the modulus values from the two rounds (Table III). Furthermore, the CV of repetitions (8.6–16%) within the rounds was notably lower compared to the CV between the rounds (37.0–38.5%). The instantaneous modulus (E_{Lab}) decreased with the histology-based ICRS scores [Fig. 3(A)]. However, this trend was not observed when the modulus ($E_{Artscan}$) and the average multimodal score of surgeons [Fig. 3(B)] were compared.

Discussion

In this study, we employed for the first time multiple arthroscopically applicable imaging and indentation methods for simultaneous evaluation of articular cartilage lesions based on the ICRS scoring system. The multimodality imaging and automatic scoring software allowed for detailed visualization and quantitative analysis of cartilage compared to conventional subjective arthroscopic evaluation. Surprisingly, no technique indicated significant superiority over the others. However, in comparison between the surgeons' and histological (gold standard) scores, the conventional arthroscope and OCT were found to be the most reliable techniques, whereas US showed the lowest reliability. Furthermore, automatic scoring software was found to be as reliable as the OCT-based scoring by the surgeons.

The average inter-method agreements suggest that conventional arthroscopic examination and OCT are superior to US. The higher resolution of OCT provides enhanced visualization of surface fibrillation compared to US. However, both OCT and US are limited by their 2D cross-sectional imaging, whereas the arthroscopic examination enables visualization of the whole AI without manual

Table I
Average ICRS scores ($N = 43$) and instantaneous moduli ($E_{Artscan}$) for both surgeons, including SD and 95% CI of each round. Corresponding gold standard values for average histology-based ICRS score and laboratory mechanical testing system based instantaneous modulus (E_{Lab}) are presented

	Surgeon 1		Surgeon 2	
	Round 1	Round 2	Round 1	Round 2
	Mean ± SD (95% CI)	Mean ± SD (95% CI)	Mean ± SD (95% CI)	Mean ± SD (95% CI)
ICRS				
Arthroscope	0.65 ± 0.13 (0.39, 0.91)	0.86 ± 0.16 (0.55, 1.17)	0.79 ± 0.14 (0.51, 1.07)	1.02 ± 0.14 (0.74, 1.30)
US	0.63 ± 0.14 (0.36, 0.90)	0.84 ± 0.16 (0.52, 1.16)	0.72 ± 0.13 (0.47, 0.98)	0.93 ± 0.13 (0.68, 1.18)
OCT	0.65 ± 0.13 (0.39, 0.91)	0.84 ± 0.16 (0.53, 1.14)	0.91 ± 0.14 (0.62, 1.19)	1.28 ± 0.13 (1.03, 1.53)
Software	0.86 ± 0.14 (0.59, 1.13)	0.91 ± 0.14 (0.62, 1.19)	1.05 ± 0.14 (0.76, 1.33)	1.33 ± 0.12 (1.08, 1.57)
Histology	1.14 ± 0.05 (1.04, 1.24)			
Biomechanics				
$E_{Artscan}$ (MPa)	7.95 ± 1.84 (4.34, 11.56)	3.06 ± 0.29 (2.49, 3.62)	5.86 ± 0.46 (4.95, 6.76)	2.78 ± 0.21 (2.37, 3.20)
E_{Lab} (MPa)	4.63 ± 0.56 (3.53, 5.73)			

Table II
Inter-method agreement and reliability relative to histological scores for both surgeons

	Round 1			Round 2		
	Agreement (%)	κ (95% CI)	*P*-value	Agreement (%)	κ (95% CI)	*P*-value
Surgeon 1						
Arthroscope	46.5	0.240 (0.040, 0.439)	0.007	55.8	0.382 (0.192, 0.572)	<0.001
US	44.2	0.211 (0.025, 0.397)	0.014	51.2	0.321 (0.131, 0.511)	<0.001
OCT	48.8	0.273 (0.077, 0.469)	0.002	48.8	0.283 (0.094, 0.472)	0.001
Software	51.2	0.304 (0.116, 0.493)	0.001	58.1	0.407 (0.210, 0.603)	<0.001
Surgeon 2						
Arthroscope	58.1	0.408 (0.207, 0.609)	<0.001	60.5	0.444 (0.241, 0.647)	<0.001
US	48.8	0.273 (0.069, 0.477)	0.002	58.1	0.404 (0.193, 0.615)	<0.001
OCT	55.8	0.377 (0.163, 0.590)	<0.001	67.4	0.540 (0.342, 0.739)	<0.001
Software	53.5	0.338 (0.129, 0.548)	<0.001	51.2	0.308 (0.106, 0.510)	0.001

Table III
Intraobserver and interobserver agreements and reproducibilities for all applied techniques

Intraobserver	Surgeon 1				Surgeon 2			
	Agreement (%)	κ (95% CI)		*P*-value	Agreement(%)	κ (95% CI)		*P*-value
Arthroscope	72.1	0.557 (0.368, 0.746)		<0.001	65.1	0.500 (0.307, 0.693)		<0.001
US	79.1	0.660 (0.494, 0.826)		<0.001	58.1	0.332 (0.097, 0.566)		0.002
OCT	76.7	0.628 (0.448, 0.809)		<0.001	53.5	0.353 (0.149, 0.558)		<0.001
Software	76.7	0.630 (0.435, 0.826)		<0.001	53.5	0.321 (0.114, 0.527)		0.001
Biomechanics	R^2 (%)	CV_{RMS} (%)	$CV_{RMS, R1}$ (%)	$CV_{RMS, R2}$ (%)	R^2 (%)	CV_{RMS} (%)	$CV_{RMS, R1}$ (%)	$CV_{RMS, R2}$ (%)
Artscan	63.6	38.5	16.0	10.3	63.3	37.0	10.1	8.6
Interobserver	Round 1				Round 2			
	Agreement (%)	κ (95% CI)		*P*-value	Agreement (%)	κ (95% CI)		*P*-value
Arthroscope	69.8	0.515 (0.311, 0.718)		<0.001	69.8	0.580 (0.403, 0.757)		<0.001
US	76.7	0.615 (0.430, 0.801)		<0.001	60.5	0.459 (0.283, 0.635)		<0.001
OCT	72.1	0.571 (0.367, 0.775)		<0.001	51.2	0.359 (0.183, 0.535)		<0.001
Software	72.1	0.582 (0.393, 0.770)		<0.001	46.5	0.237 (0.051, 0.423)		0.011

*CV_{RMS} = Coefficient of variation between two measurement rounds, $CV_{RMS, R1/R2}$ = Coefficient of variation between the repetitions within the first (R1) and second (R2) rounds separately.

mapping. Nevertheless, the axial scan function of the OCT and US imaging systems enables fast imaging of relatively large areas, thus supporting especially OCT's superiority over the other techniques when combined with the automatic scoring software. Overall, the inter-method reliability (Table II) between the imaging techniques and the histology were lower compared to previous studies[23,31]. The systematically lower surgeons' scores with these techniques, especially with ICRS 1, further emphasize the inferior resolution of the techniques compared to histology and the importance of the detection of the surface fibrillation. Furthermore, the inter-method agreements were relatively lower compared to intraobserver and interobserver agreements. This finding was to be expected as the methods assess various aspects of cartilage matrix at different resolutions, thus suggesting their complementarity in the evaluation of cartilage degeneration. Moreover, this finding highlights the importance of accurate site matching between the techniques, which was hindered by the surgeons' ability to distinguish the most severe lesion within an AI.

Surgeon 1 systematically assigned lower scores in the first round for the same cartilage lesions, whereas surgeon 2 was likely to find more advanced cartilage lesions in the second round (Table I). This observation is consistent with previous studies where it was concluded that surgeon-based scoring can be subjective and poorly reliable, even with imaging-based techniques[8–10]. Moreover, the

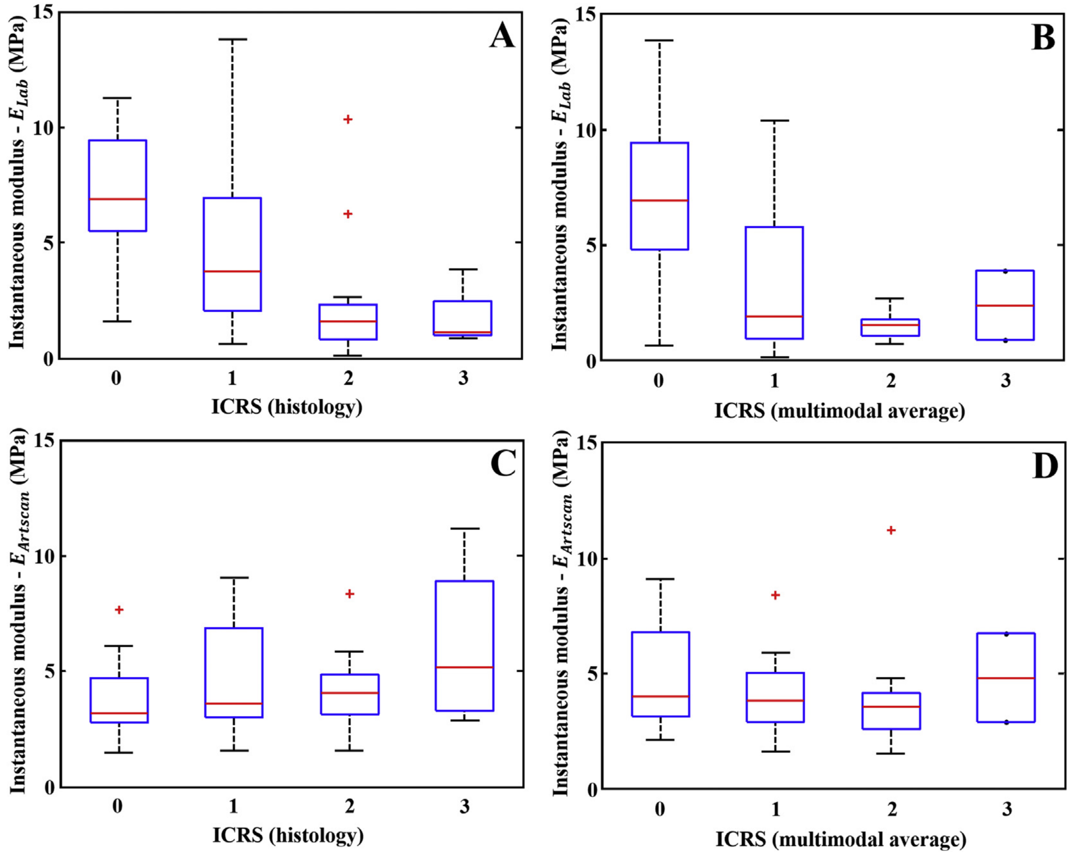

Fig. 3. Biomechanical response of equine cartilage as a function of ICRS score. (A) An expected decrease of cartilage stiffness (E_{Lab}) is observed with higher ICRS score (histology). (B) Notable difference in cartilage stiffness (E_{Lab}) is observed between ICRS grades 0 and 1 based on the average score of multimodal scorings. (C–D) A similar trend is not apparent with Artscan measurements ($E_{Artscan}$) and ICRS scores (histology) or the average score of multimodal scorings. A single measurement point is not visible (27.7 MPa, at ICRS 1 and ICRS 0) in subfigures C and D, respectively.

reproducibility of scoring was found being governed by the surgeons' ability to reproducibly identify the most severe cartilage lesion within an AI, hence suggesting a necessity for imaging the articular surface in its entirety[34] and for automatic quantitative scoring of all chondral lesions[11,12]. Related to a clinical situation, implementation of these protocols would improve the reliability and level of details in the cartilage evaluation; however, the joint morphology limits access for the evaluation of the whole articular surface. Scores based on OCT images were observed to be generally higher compared to those based on arthroscope or US (Table I), arguably due to the higher resolution of OCT[22], thus emphasizing the distinction of cartilage surface fibrillation (ICRS 1) from the OCT images. The average histological score of independent evaluators was found to be relatively higher compared to the surgeons' scores, except for the average score of surgeon 2 on the second round with OCT. The superior resolution of histological images enables enhanced visualization of cartilage fibrillation, hence resulting in higher frequency of ICRS 1. Similar findings have been reported earlier[31,34].

The average ICRS scores determined using arthroscope or based on US and OCT images increased and average cartilage stiffness decreased with both surgeons on the second round (Table I), hence indicating that lesion finding improved on the second round. However, this might also indicate that the cartilage surface was damaged during the first evaluation round. As no contact was established between the imaging devices and the cartilage surface, the possible damage may have occurred during the arthroscopic indentation measurement. Nevertheless, great care was taken during the measurements to avoid damaging the cartilage surface adjacent to the measurement location. Furthermore, as the size of the indenter is small compared to the size of the AI, measurement of the same exact location is highly improbable. The lower SD values of cartilage stiffness measurement ($E_{Artscan}$) on the second round indicate increased reproducibility of the measurements, as observed from the CV values (Table III), induced by the increased experience of surgeons as the measurements progressed. This may also explain the systematic difference in the ICRS scores between the rounds.

The intraobserver and interobserver reproducibility values via conventional arthroscope presented in this study (Table III) are superior[9] or similar[10,23] with the previous findings. Reproducibility of US-based ICRS scoring has not been previously reported, whereas the reproducibility of OCT-based scoring obtained here were slightly lower than that reported in an earlier study[23,31]. This may be because the reported reproducibilities of previous studies[9,10,23,31] were performed by scoring identical images in multiple repetitions; however, here the imaging and scoring of the AIs were, for the first time, repeated by both surgeons on both rounds. Interestingly, the interobserver agreements decreased on the second round, apart from conventional arthroscopic scoring. By comparing the average ICRS scores between the rounds (Table I), the decrease of US scoring reproducibility was unexpected and probably results from the difficulty to distinguish between ICRS 0–2 grades as also observed from the lowest inter-method agreement and reliability of the technique. However, for OCT, and thus automatic software, the decrease results from the surgeon 2 finding more severe chondral lesions on the second round. Additionally, high resolution OCT allowed imaging of small chondral flaps which were not visible with US or through conventional arthroscopic view, hence limiting the assessment of ICRS 1 lesions.

Automated software systematically assigned higher ICRS scores for the lesions in OCT images than the surgeons in the two rounds (Table I). The thresholds for assigning ICRS scores were optimized with the data of previous studies[23,31], hence indicating the surgeons' scores to be relatively lower compared to the evaluators in previous studies. In the OCT-based scoring, the agreement and reproducibility of software were similar to that of surgeons (Table III). Similar results have been reported earlier[31]. Semi-automatic software approaches were recently proposed for cartilage grading[11,12]; however, these solutions still depend on some degree of user input. In contrast to earlier laboratory studies, the software scoring algorithm applied in this study was optimized for cartilage images obtained with an intravascular OCT catheter, which can be easily applied during arthroscopy[17]. The thresholds applied in the automatic algorithm were based on ORI and relative lesion depth obtained using a relatively large ($N = 148$) independent image and scoring sets[23,31]. Nevertheless, we believe that the accuracy and sensitivity of the algorithm can be further improved by using an even larger training set of OCT images.

The scores of healthy (ICRS 0) and highly damaged cartilage (ICRS 4) are the easiest to classify[8–10]; hence, the high resolution OCT was chosen as the development platform for the algorithm to reliably differentiate the ICRS scores 1–3. The drawback of OCT compared to US is the inferior contrast of cartilage–bone interface, therefore superimposing of US and OCT images could enable more reliable quantitative analysis. However, the superposition of images was not attempted as perfect orientation is required with both imaging catheters. Furthermore, the resolution difference of the imaging methods must be taken into account. For a clinical situation, the axial scanning function of the imaging catheters enables fast imaging of relatively large areas, thus making the both techniques feasible in the clinical time constraints. However, the automatic software must be further optimized from 2D to 3D images to facilitate fast *in vivo* analyses.

Chondral softening may indicate compromised cartilage competence, even if the cartilage appears visually intact under an arthroscopic examination. Cartilage stiffness (E_{Lab}) exhibited a decreasing trend as the histology-based ICRS scores of the AIs increased, which is consistent with earlier findings[18,35]. The CV of measurements with the arthroscopic indentation device (Artscan) (8.6–16%) with subsequent repetitions is similar to an earlier study[21]. However, a similar decreasing trend was not observed with instantaneous moduli ($E_{Artscan}$) and the histological and average multimodal scorings, possibly due to the site-dependent variation in cartilage stiffness[27–29] within each AI, and the surgeons' inability to locate and score the exact same lesion site in the subsequent round.

Location matching between the imaging techniques and histology resulted in a single outlier which was excluded from the study. Also, the determination of cartilage stiffness with the arthroscopic device (Artscan 200) was found to be time-consuming and prone to errors related to the need for perpendicular alignment of the indenter against the articular surface as observed by the high CV between the rounds. Based on the recent studies a rapid non-destructive method, such as NIR spectroscopy[15,19], could be one feasible solution for estimation of cartilage mechanical properties.

In conclusion, neither of the hypotheses could be explicitly proven. OCT imaging combined with the automatic scoring software provided the most detailed and reliable scoring of chondral lesions. Furthermore, arthroscopic examination was found as reliable as visual scoring of lesions in OCT images, whereas US had the lowest reliability. However, the limited penetration depth of light limits the clinical potential of OCT in human cartilage thickness assessment, thus implying the combination of OCT and US could be optimal solution for reliable diagnostics. The automatic software was only as reliable as surgeon-based scoring, thus implying the main factor affecting the reliability was difficulty in locating the same lesion in the separate imaging sessions. These findings suggest that cartilage area should be imaged and quantitatively analyzed in its entirety to eliminate surgeon-related bias, therefore providing superior reliability over conventional arthroscopy.

Contributions

JK Sarin: contributed in equine *ex vivo* study (data acquisition), data analysis, interpretation of the data and was the main writer of the manuscript.

H Brommer and **D Argüelles**: contributed in equine *ex vivo* study (arthroscopies and scoring).

PH Puhakka and **SI Inkinen**: contributed in equine *ex vivo* study (data acquisition).

IO Afara: contributed in the study design, equine *ex vivo* study and interpretation of data.

S Saarakkala: contributed in preparation of histological images, and histological scoring.

J Töyräs: contributed in the study design, equine *ex vivo* study, histological scoring and interpretation of data.

All authors contributed in the preparation and approval of the final submitted manuscript.

Competing interests

The authors have no conflicts of interest related to the execution of this study and preparation of the manuscript.

Role of funding

This study was financially supported by Doctoral Programme in Science, Technology and Computing (SCITECO) of University of Eastern Finland, the strategic funding of the University of Eastern Finland and University of Oulu, the Academy of Finland (project 267551, University of Eastern Finland and project 268378, University of Oulu) and Kuopio University Hospital (VTR projects 5041750 and 5041744, PY210 Clinical Neurophysiology).

Acknowledgements

We acknowledge PhD Virpi Tiitu for her assistance in histological scoring and MSc Mikko Pitkänen for his assistance in the measurements.

References

1. Moores AP, Benigni L, Lamb CR. Computed tomography versus arthroscopy for detection of canine elbow dysplasia lesions. Vet Surg 2008;37(4):390–8, http://dx.doi.org/10.1111/j.1532-950X.2008.00393.x.
2. Von Engelhardt LV, Lahner M, Klussmann A, Bouillon B, Dàvid A, Haage P, *et al*. Arthroscopy vs. MRI for a detailed assessment of cartilage disease in osteoarthritis: diagnostic value of MRI in clinical practice. BMC Musculoskelet Disord 2010;11(1):75, http://dx.doi.org/10.1186/1471-2474-11-75.
3. Dahill M, Stevenson AJ, Hughes AM, Williams JL. Comparison of arthroscopic and MRI findings of osteochondral damage in knees. Bull Hosp Joint Dis 2014;72(4):284–7.
4. Kijowski R, Blankenbaker DG, Stanton PT, Fine JP, De Smet AA. Radiographic findings of osteoarthritis versus arthroscopic findings of articular cartilage degeneration in the tibiofemoral

joint. Radiology 2006;239(3):818–24, http://dx.doi.org/10.1148/radiol.2393050584.
5. Brittberg M, Winalski CS. Evaluation of cartilage injuries and repair. J Bone Joint Surg Am 2003;85-A(Suppl 2):58–69.
6. Outerbridge R. The etiology of chondromalacia patellae. J Bone Joint Surg Br 1961;43-B:752–7.
7. Dougados M, Ayral X, Listrat V, Gueguen A, Bahuaud J, Beaufils P, *et al*. The SFA system for assessing articular cartilage lesions at arthroscopy of the knee. Arthroscopy 1994;10(1): 69–77, http://dx.doi.org/10.1016/S0749-8063(05)80295-6.
8. Spahn G, Klinger HM, Hofmann GO. How valid is the arthroscopic diagnosis of cartilage lesions? Results of an opinion survey among highly experienced arthroscopic surgeons. Arch Orthop Trauma Surg 2009;129(8):1117–21, http://dx.doi.org/10.1007/s00402-009-0868-y.
9. Spahn G, Klinger HM, Baums M, Pinkepank U, Hofmann GO. Reliability in arthroscopic grading of cartilage lesions: results of a prospective blinded study for evaluation of inter-observer reliability. Arch Orthop Trauma Surg 2011;131(3):377–81, http://dx.doi.org/10.1007/s00402-011-1259-8.
10. Brismar BH, Wredmark T, Movin T, Leandersson J, Svensson O. Observer reliability in the arthroscopic classification of osteoarthritis of the knee. J Bone Joint Surg Br 2002;84(1):42–7, http://dx.doi.org/10.1302/0301-620X.84B1.11660.
11. Cernohorsky P, Kok AC, Bruin DM de, Brandt MJ, Faber DJ, Tuijthof GJ, *et al*. Comparison of optical coherence tomography and histopathology in quantitative assessment of goat talus articular cartilage. Acta Orthop 2015;86(2):257–63, http://dx.doi.org/10.3109/17453674.2014.979312.
12. Nebelung S, Marx U, Brill N, Arbab D, Quack V, Jahr H, *et al*. Morphometric grading of osteoarthritis by optical coherence tomography – an ex vivo study. J Orthop Res 2014;32(10): 1381–8, http://dx.doi.org/10.1002/jor.22673.
13. Liukkonen J, Hirvasniemi J, Joukainen A, Penttilä P, Virén T, Saarakkala S, *et al*. Arthroscopic ultrasound technique for simultaneous quantitative assessment of articular cartilage and subchondral bone: an in vitro and in vivo feasibility study. Ultrasound Med Biol 2013;39(8):1460–8, http://dx.doi.org/10.1016/j.ultrasmedbio.2013.03.026.
14. Fujimoto JG, Pitris C, Boppart SA, Brezinski ME. Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia 2000;2(1–2):9–25, http://dx.doi.org/10.1038/sj.neo.7900071.
15. Afara IO, Hauta-Kasari M, Jurvelin JS, Oloyede A, Töyräs J. Optical absorption spectra of human articular cartilage correlate with biomechanical properties, histological score and biochemical composition. Physiol Meas 2015;36(9):1913–28, http://dx.doi.org/10.1088/0967-3334/36/9/1913.
16. Huang Y-P, Zheng Y-P. Intravascular ultrasound (IVUS): a potential arthroscopic tool for quantitative assessment of articular cartilage. Open Biomed Eng J 2009;3:13–20, http://dx.doi.org/10.2174/1874120700903010013.
17. Li X, Martin S, Pitris C, Ghanta R, Stamper DL, Harman M, *et al*. High-resolution optical coherence tomographic imaging of osteoarthritic cartilage during open knee surgery. Arthritis Res Ther 2005;7(2):318–23, http://dx.doi.org/10.1186/ar1491.
18. Marticke JK, Hösselbarth A, Hoffmeier KL, Marintschev I, Otto S, Lange M, *et al*. How do visual, spectroscopic and biomechanical changes of cartilage correlate in osteoarthritic knee joints? Clin Biomech (Bristol, Avon) 2010;25(4):332–40, http://dx.doi.org/10.1016/j.clinbiomech.2009.12.008.
19. Sarin JK, Amissah M, Brommer H, Argüelles D, Töyräs J, Afara IO. Near infrared spectroscopic mapping of functional properties of equine articular cartilage. Ann Biomed Eng 2016;44(11):3335–45, http://dx.doi.org/10.1007/s10439-016-1659-6.
20. Lyyra T, Jurvelin J, Pitkänen P, Väätäinen U, Kiviranta I. Indentation instrument for the measurement of cartilage stiffness under arthroscopic control. Med Eng Phys 1995;17(5):395–9, http://dx.doi.org/10.1016/1350-4533(95)97322-G.
21. Brommer H, Laasanen MS, Brama PAJ, van Weeren PR, Helminen HJ, Jurvelin JS. In situ and ex vivo evaluation of an arthroscopic indentation instrument to estimate the health status of articular cartilage in the equine metacarpophalangeal joint. Vet Surg 2006;35(3):259–66, http://dx.doi.org/10.1111/j.1532-950X.2006.00136.x.
22. Virén T, Huang YP, Saarakkala S, Pulkkinen H, Tiitu V, Linjama A, *et al*. Comparison of ultrasound and optical coherence tomography techniques for evaluation of integrity of spontaneously repaired horse cartilage. J Med Eng Technol 2012;36(3):185–92, http://dx.doi.org/10.3109/03091902.2012.663054.
23. Niemelä T, Virén T, Liukkonen J, Argüelles D, te Moller NCR, Puhakka PH, *et al*. Application of optical coherence tomography enhances reproducibility of arthroscopic evaluation of equine joints. Acta Vet Scand 2014;56(1):3, http://dx.doi.org/10.1186/1751-0147-56-3.
24. Appleyard RC, Swain MV, Khanna S, Murrell GA. The accuracy and reliability of a novel handheld dynamic indentation probe for analysing articular cartilage. Phys Med Biol 2001;46(2): 541–50, http://dx.doi.org/10.1088/0031-9155/46/2/319.
25. Treppo S, Koepp H, Quan EC, Cole AA, Kuettner KE, Grodzinsky AJ. Comparison of biomechanical and biochemical properties of cartilage from human knee and ankle pairs. J Orthop Res 2000;18(5):739–48, http://dx.doi.org/10.1002/jor.1100180510.
26. Julkunen P, Harjula T, Iivarinen J, Marjanen J, Seppänen K, Närhi T, *et al*. Biomechanical, biochemical and structural correlations in immature and mature rabbit articular cartilage. Osteoarthritis Cartilage 2009;17(12):1628–38, http://dx.doi.org/10.1016/j.joca.2009.07.002.
27. Athanasiou KA, Rosenwasser MP, Buckwalter JA, Malinin TI, Mow VC. Interspecies comparisons of in situ intrinsic mechanical properties of distal femoral cartilage. J Orthop Res 1991;9(3):330–40, http://dx.doi.org/10.1002/jor.1100090304.
28. Wang S-Z, Huang Y-P, Saarakkala S, Zheng Y-P. Quantitative assessment of articular cartilage with morphologic, acoustic and mechanical properties obtained using high-frequency ultrasound. Ultrasound Med Biol 2010;36(3):512–27, http://dx.doi.org/10.1016/j.ultrasmedbio.2009.12.005.
29. Mäkelä JTA, Rezaeian ZS, Mikkonen S, Madden R, Han S-K, Jurvelin JS, *et al*. Site-dependent changes in structure and function of lapine articular cartilage 4 weeks after anterior cruciate ligament transection. Osteoarthritis Cartilage 2014;22(6):869–78, http://dx.doi.org/10.1016/j.joca.2014.04.010.
30. Hayes WC, Keer LM, Herrmann G, Mockros LF. A mathematical analysis for indentation tests of articular cartilage. J Biomech 1972;5(5):541–51, http://dx.doi.org/10.1016/0021-9290(72)90010-3.
31. Te Moller NCR, Pitkänen M, Sarin JK, Väänänen S, Liukkonen J, Afara IO, *et al*. Semi-automated ICRS scoring of equine articular cartilage lesions in optical coherence tomography images. Equine Vet J 2016, http://dx.doi.org/10.1111/evj.12637.
32. Saarakkala S, Wang S-Z, Huang Y-P, Zheng Y-P. Quantification of the optical surface reflection and surface roughness of articular cartilage using optical coherence tomography. Phys

Med Biol 2009;54(22):6837–52, http://dx.doi.org/10.1088/0031-9155/54/22/006.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74, http://dx.doi.org/10.2307/2529310.
34. Nebelung S, Brill N, Marx U, Quack V, Tingart M, Schmitt R, *et al*. Three-dimensional imaging and analysis of human cartilage degeneration using optical coherence tomography. J Orthop Res 2015;33(5):651–9, http://dx.doi.org/10.1002/jor.22828.
35. Kleemann RU, Krocker D, Cedraro A, Tuischer J, Duda GN. Altered cartilage mechanics and histology in knee osteoarthritis: relation to clinical assessment (ICRS grade). Osteoarthritis Cartilage 2005;13(11):958–63, http://dx.doi.org/10.1016/j.joca.2005.06.008.