RA [1, 2]. The international standard for performing RA US scanning and evaluation of disease activity is the OMERACT-ELUAR Synovitis Scoring (OESS) system [1, 3].

To further mitigate the operator-dependency in scoring disease activity on CD US images in future trials and clinical practice, we proposed the use of convolutional neural networks (CNN) to automatically grade CD US images according to the OESS definitions. This study is a continuation of the findings in our previous work, where we developed a CNN for four-class CD US OESS scoring with a test accuracy of 75.0% [4].

Objectives: Since our last contribution, we have further developed the architecture of this neural network and can here present a new idea applying a Cascaded Convolutional Neural Network design. We evaluate the generalizability of this method on unseen data, comparing the CNN with an expert rheumatologist.

Methods: The images used for developing the algorithms were graded by a single expert rheumatologist according to the OESS system. The CNNs in the cascade were trained individually, after which they were combined to form the cascade model as shown in figure 1. The algorithms were evaluated on a separate test dataset, which came from the same distribution as the training dataset. The algorithms were compared to the gradings of an expert rheumatologist on a per-joint basis using a Kappa test, and on a per-patient basis using a Wilcoxon Signed Rank test.

Figure 1. CNN-1 is the first CNN in the model and distinguishes between RA disease grade (DG) 0 and DG's 1, 2 and 3. CNN-2 is the second CNN and distinguishes between DG 1 and DG's 2 and 3. CNN-3 is the final CNN which distinguishes between DG's 2 and 3.

Results: With 1678 images available for training and 322 images for testing the model, the model achieved an overall 4-class accuracy of 83.9%. On a per-patient level, there was no significant difference between the classifications of the model and of a human expert ($p=0.85$).

Conclusion: We have shown that dividing a four-degree classification task into three successive binary classification tasks has resulted in a model capable of making correct classifications in 83.9% of the cases for a test set of ultrasound images with a naturally occurring distribution of RA joint disease activity scores. Furthermore, we have shown that the cascade model can produce classification decisions comparable with a human rheumatologist when applied on a per-patient basis. This emphasizes the potential of using CNNs with this architecture as a strong assistive tool for the objective assessment of disease activity of RA patients.

References:

- 1) D'Agostino M-A, Terslev L, Aegerter P, et al. (2017). Scoring ultrasound synovitis in rheumatoid arthritis: a OMERACT-ELUAR ultrasound taskforce-Part 1: definition and development of a standardised, consensus-based scoring system. RMD Open.
- 2) Paulsush NS, Aga A-B, Olsen I, et al. (2018). Clinical and ultrasound remission after 6 months of treat-to-target therapy in early rheumatoid arthritis: associations to future good radiographic and physical outcomes. Ann Rheum Dis, 77, s. 1425-25.
- 3) Terslev L, Naredo E, Aegerter P, et al. (3 2017). Scoring ultrasound synovitis in rheumatoid arthritis: a OMERACT-ELUAR ultrasound taskforce-Part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. RMD Open.
- 4) Andersen JKH, Pedersen JS, Laursen MS, et al. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. RMD Open 2019; 5:e000891. doi:10.1136/rmdopen-2018-000891

Disclosure of Interests: None declared

DOI: 10.1136/annrheumdis-2020-eular.888

OP0065

INTER-RATER RELIABILITY AND VALIDITY OF MEDIAL AND LATERAL FEMORAL BONE MARROW LESIONS IN PATIENTS WITH OSTEOARTHRITIS UNDERGOING TIBIAL OSTEOTOMY USING THE KNEE INFLAMMATION MRI SCORING SYSTEM (KIMRISS)

J. Schulz^{1,2}, T. Birmingham^{1,2}, W. P. Maksymowych³, R. G. Lambert³, S. Pritchett⁴, F. Beier^{2,5}, J. R. Giffin^{2,6,7}, T. Appleton^{2,5,6}. ¹Western University, Health and Rehabilitation Sciences, London, Canada; ²Western University, Bone and Joint Institute, London, Canada; ³University of Alberta, Medicine, Edmonton, Canada; ⁴London Health Sciences Centre, Radiology, London,

Canada; ⁵Western University, Physiology and Pharmacology, London, Canada;

⁶Western University, Medicine, London, Canada; ⁷London Health Sciences Centre, Surgery, London, Canada

Background: The Knee Inflammation MRI Scoring System (KIMRISS) is a semi-quantitative grading tool (range 0-500) used to measure features of inflammation in knee osteoarthritis (OA), including compartment-specific BML scores.

Objectives: To investigate; 1) Inter-rater reliability of KIMRISS BML change scores including expert readers and a trainee reader, 2) Validity of KIMRISS BML scores by comparing medial and lateral femoral compartments in surgical and contralateral limbs before and 1 year after medial opening wedge high tibial osteotomy (HTO).

Methods: We included 27 patients with varus alignment and medial compartment knee OA [preoperative Kellgren-Lawrence grades ≤ 3 , age = 53.1 ± 5.9 , 22 males (81%), body mass index = 29.9 ± 4.3] undergoing unilateral HTO. Sagittal 2D turbo spin-echo sequences were acquired at 3-Tesla in both knees pre and 1-year post HTO after surgical removal of hardware. Femoral BML scores in the medial and lateral compartments were assessed on images blinded to limb and time by masking surgical evidence. Three reviewers independently graded the same 108 images (27 patients, 2 limbs, 2 time points). We evaluated reliability pre-to-post HTO BML change scores in the medial and lateral femur by calculating intraclass correlation coefficients (ICC) and Bland-Altman plots with 80% Limits of Agreement (LoA) for each rater pair.

Results: Means and standard deviations are shown in Figure 1. Cumulative percent distribution plots of changes after surgery show both increases and decreases in the medial and lateral femur (Figure 2). When combining all raters, the ICC (95% confidence interval (CI)) for the change in BML score for the surgical knee was 0.81 (0.64, 0.91) in the medial femur and 0.73 (0.49, 0.87) in the lateral femur. 80% LOA for change in BML score in the surgical knee was -3.56 to 5.44. ICCs and LoAs for rater pairs are in Table 1. The surgical limb medial femur BML score was significantly greater than all other scores pre and post HTO (Figure 1). There were no significant changes after HTO. The mean change (95% CI) in medial femur BML score of the surgical knee was 1.08 (-1.21, 3.38). Using the mean scores of all raters, we compared BML scores in surgical and contralateral limbs before and after HTO using a 2-way repeated measures (limb by time) ANOVA, separately for the medial and lateral compartments.

Table 1. Intraclass Correlation Coefficients (95% Confidence Intervals) and 80% Limits of Agreement for change in BMLs scores in the surgical knee.

a) Medial compartment			
	1	2	3
1	0.63 (0.34, 0.81)		
2			
3	0.69 (0.43, 0.85)	0.42 (0.06, 0.69)	
b) Lateral compartment			
	1	2	3
1			
2	0.67 (0.41, 0.83)		
3	0.79 (0.60, 0.90)	0.47 (0.11, 0.72)	
c) Medial compartment			
	1	2	3
1			
3	-6.09, 8.61		
4	-8.10, 9.65	-6.03, 7.58	
d) Lateral compartment			
	1	2	3
1			
2	-3.49, 5.41		
3	-4.73, 6.99	-3.43, 4.95	

ICCs in the medial (a) and lateral (b) compartments, and 80% LoA in the medial (c) and lateral (d) compartments. Rater 1 = trainee reader, raters 2-3 = expert readers (one MSK radiologist and one rheumatologist).

Conclusion: The KIMRISS can reliably detect differences between femoral BML scores in symptomatic and contralateral limbs, supporting the inter-rater reliability, feasibility and validity of compartment-specific BML scores.

Disclosure of Interests: Jenna Schulz: None declared, Trevor Birmingham: None declared, Walter P Maksymowych Grant/research support from: Received research and/or educational grants from Abbvie, Novartis, Pfizer,

Fig 1. Means and standard deviations for KIMRISS femoral BML scores in the medial and lateral compartment for the surgical and contralateral knee.

Fig 2. Probability plots for the 12-month change after surgery in femoral BML score in the medial (a) and lateral (b) compartments of the surgical knee. Change scores from each of the three raters is shown, as well as the average of all four raters.

UCB, Consultant of: WPM is Chief Medical Officer of CARE Arthritis Limited, has received consultant/participated in advisory boards for Abbvie, Boehringer, Celgene, Eli-Lilly, Galapagos, Gilead, Janssen, Novartis, Pfizer, UCB, Speakers bureau: Received speaker fees from Abbvie, Janssen, Novartis, Pfizer, UCB., Robert G Lambert: None declared, Stephany Pritchett: None declared, Frank

Beier: None declared, J. Robert Giffin: None declared, Thomas Appleton Grant/ research support from: AbbVie and Pfizer

DOI: 10.1136/annrheumdis-2020-eular.2002

OP0066

IMAGING-DETECTED FEATURES OF HAND OSTEOARTHRITIS ASSOCIATE WITH SYMPTOMS AND RADIOGRAPHIC CHANGE OVER TIME: A SYSTEMATIC REVIEW AND META-ANALYSIS OF OBSERVATIONAL STUDIES

A. Obotiba¹, S. Swain¹, J. Kaur¹, K. Yaseen¹, M. Doherty¹, W. Zhang¹, A. Abhishek¹. ¹University of Nottingham, Academic Rheumatology, Division of Rheumatology, Orthopaedics and Dermatology, Nottingham, United Kingdom

Background: Osteoarthritis (OA) commonly affects joints in the hand. The natural history of hand OA is not well understood, and the local determinants of symptoms and structural changes over time remain unclear.

Objectives: To investigate, in both cross-sectional and prospective studies, the association between imaging (ultrasound [US] and magnetic resonance imaging [MRI]) features and symptoms of hand OA, and to examine in prospective studies whether imaging-detected features at baseline predict subsequent clinical and radiographic outcomes.

Methods: A systematic literature search was conducted in five databases including Medline, Web of Science, EMBASE, CINAHL and AMED in April 2018. The search was designed to capture published observational studies on the use of US and MRI in hand OA with no language restrictions. Odds ratios (OR), risk ratios (RR), and 95% confidence interval (CI) between [1] imaging features and hand OA symptoms at baseline, and [2] baseline-imaging features and follow-up outcomes were extracted and pooled using random effects model. Outcomes were defined as either incidence or progression of pre-existing features. Risk of bias assessment was performed using the Newcastle-Ottawa Scales. Heterogeneity and publication bias were assessed.

Results: The search identified 2818 citations, which reduced to 2216 after duplicate removal. Screening of titles and abstracts found 140 articles which met the inclusion criteria. After full text screening, 25 were included for analysis, including 452 participants (87% women) for US and 298 participants (86% women) for MRI with mean ages 60.3 and 62.5, respectively. Imaging-detected structural OA features were preferentially found in distal interphalangeal joints (DIPJs) followed by carpometacarpal (CMCJ) and proximal interphalangeal (PIPJ) joints. Metacarpophalangeal joints were least affected. However, the distribution pattern was different for inflammatory features for which the CMCJ was the most affected, and with no clear difference between DIPJs and PIPJs (Figure 1).

Figure 1. Hand map of grey-scale synovitis derived from pooled estimates of prevalence across studies (%[95% CI])

Of 10 US and 5 MRI studies examining association at baseline, joint tenderness was associated with US osteophytes (pooled ORs 2.30, 95% CI 1.90-2.79), grey-scale synovitis (3.00, 2.33-3.84), synovial effusion (2.92, 2.29-3.72), and power Doppler (PD) (2.30, 1.68-3.15). Similar relationships were observed with MRI features (Figure 2). Six studies did not find any association between imaging features and self-reported outcomes. However, association was observed with US- and MRI-detected synovitis in one study each, and MRI-detected structural features in two. Statistical pooling was not possible for these outcomes due to heterogeneous data.

Of the 9 US and 5 MRI studies for prediction, a dose-dependent relationship was observed between baseline PD and radiographic change at follow-up (Figure 3). Similar results were observed for MRI features and Kellgren-Lawrence change. The pooled ORs (95% CI) was 2.66 (1.88, 3.78) for bone marrow lesions, and 2.18 (1.53, 3.10) and 4.7 (3.08, 7.18) for grades 1 and 2 synovitis, respectively. Data to predict change in clinical outcomes however, were lacking.