# Validation of a Knowledge Transfer Tool for the Knee Inflammation MRI Scoring System for Bone Marrow Lesions According to the OMERACT Filter: Data from the Osteoarthritis Initiative

Jacob L. Jaremko, Omar Azmat, Robert G. Lambert, Paul Bird, Ida K. Haugen, Lennart Jans, Ulrich Weber, Naomi Winn, Veronika Zubler, and Walter P. Maksymowych

*ABSTRACT.* ***Objective.*** To assess feasibility and reliability of scoring bone marrow lesions (BML) on knee magnetic resonance imaging (MRI) in osteoarthritis using the Outcome Measures in Rheumatology Knee Inflammation MRI Scoring System (KIMRISS), with a Web-based interface and online training with real-time iterative calibration.

***Methods.*** Six readers new to the KIMRISS (3 radiologists, 3 rheumatologists) scored sagittal T2-weighted fat-saturated MRI in 20 subjects randomly selected from the Osteoarthritis Initiative data, at baseline and 1-year followup. In the KIMRISS, the reader moves a transparent overlay grid within a Web-based interface to fit bones, then clicks or touches each region containing BML per slice, to score 1 if BML is present. Regional and total scores are automatically calculated. Outcomes include the interreader intraclass correlation coefficients (ICC) and the smallest detectable change (SDC).

***Results.*** Scoring took 3–12 min per scan and all readers rated the process as moderately to very user friendly. Despite a low BML burden (average score 2.8% of maximum possible) and small changes, interobserver reliability was moderate to high for BML status and change in the femur and tibia (ICC 0.78–0.88). Four readers also scored the patella reliably, whereas 2 readers were outliers, likely because of image artifacts. SDC of 1.5–5.6 represented 0.7% of the maximum possible score.

***Conclusion.*** We confirmed feasibility of knee BML scoring by new readers using interactive training and a Web-based touch-sensitive overlay system, finding high reliability and sensitivity to change. Further work will include adjustments to training materials regarding patellar scoring, and study in therapeutic trial datasets with higher burden of BML and larger changes. (First Release April 1 2017; J Rheumatol 2017;44:1718–22; doi:10.3899/jrheum.161102)

*Key Indexing Terms:*

| | | |
|---|---|---|
| KNEE JOINT | OSTEOARTHRITIS | MRI |
| SCORING METHODS | OMERACT | BONE MARROW LESION |

With emerging knee osteoarthritis (OA) therapies, it is increasingly important to objectively quantify disease status and treatment response. Magnetic resonance imaging (MRI)–based semiquantitative scoring systems generally assess knee OA by whole-organ approach [Boston Leeds Osteoarthritis Knee Score (BLOKS)[1], MRI Osteoarthritis Knee Score (MOAKS)[2], Whole Organ magnetic Resonance iMaging Score (WORMS)[3]]. Scoring can be time-consuming, and

*From the Department of Radiology and Diagnostic Imaging, and the Division of Rheumatology, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; Division of Medicine, University of New South Wales, Sydney, Australia; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Department of Radiology and Medical Imaging, Ghent University Hospital, Ghent, Belgium; King Christian 10th Hospital for Rheumatic Diseases, Gråsten; Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; Department of Radiology, Robert Jones and Agnes Hunt Orthopaedic Hospital, Oswestry, UK; Department of Radiology, Balgrist University Hospital, Zurich, Switzerland.*

*Supported by the Capital Health Chair in Diagnostic Imaging. Dr. Jaremko is the Capital Health Chair in Diagnostic Imaging at the University of Alberta.*

*J.L. Jaremko, MD, PhD, FRCPC, Department of Radiology and Diagnostic Imaging, Faculty of Medicine and Dentistry, University of Alberta; O. Azmat, MB, FRCP, Department of Radiology and Diagnostic Imaging, Faculty of Medicine and Dentistry, University of Alberta; R.G. Lambert, MB, FRCPC, Department of Radiology and Diagnostic Imaging, Faculty of Medicine and Dentistry, University of Alberta; P. Bird, MD, Division of Medicine, University of New South Wales; I.K. Haugen, MD, PhD, Department of Rheumatology, Diakonhjemmet Hospital; L. Jans, MD, PhD, Department of Radiology and Medical Imaging, Ghent University Hospital; U. Weber, MD, King Christian 10th Hospital for Rheumatic Diseases, and Institute of Regional Health Research, University of Southern Denmark; N. Winn, MBBS, FRCR, Department of Radiology, Robert Jones and Agnes Hunt Orthopaedic Hospital; V. Zubler, MD, Department of Radiology, Balgrist University Hospital; W.P. Maksymowych, MB ChB, FRCP(C), FACP, Division of Rheumatology, Faculty of Medicine and Dentistry, University of Alberta.*

*Address correspondence to Dr. J.L. Jaremko, Radiologist and Assistant Professor, Department of Radiology and Diagnostic Imaging, Faculty of Medicine, University of Alberta, 2A2.41 WMC, 8440 – 112 St. NW, Edmonton, Alberta T6G 2B7, Canada. E-mail: jjaremko@ualberta.ca*

*Accepted for publication February 14, 2017.*

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.



Downloaded on November 11, 2026 from www.jrheum.org

features of structural damage may change little during clinical trials, where prognosis may relate more to active disease involving bone marrow lesions (BML)[4]. We developed the Outcome Measures in Rheumatology (OMERACT) Knee Inflammation MRI Scoring System (KIMRISS) using a wide scoring range to optimize sensitivity to change, binary scoring to improve reliability, and an integrated standardized new-reader calibration method to optimize feasibility.

BML is typically scored separately for multiple periarticular bone regions. KIMRISS scoring is binary: Is BML present? Yes/No (1/0). The BLOKS/WORMS/MOAKS[1,2,3] involve more complex decisions, estimating percentages of fewer large regions containing BML. New readers train by reading published manuscripts and performing in-person exercises with experienced readers. In contrast, we developed a novel approach allowing systematic reader calibration without direct expert supervision. In this real-time iterative calibration (RETIC), readers use a Web-based digital overlay superimposing the outline of scoring regions on MRI. As the reader completes the scoring of each case, each overlay region changes color, indicating concordance or discordance with expert scores. This interactive feedback allows new users to rapidly align their scoring with experts as they progress through cases.

We tested the feasibility and reliability of the KIMRISS knee OA BML scoring using Web-based digital overlay and RETIC technology. We applied the relevant aspects of the OMERACT Filter 2.0[5,6]. Members of a subgroup of the OMERACT MRI in the Arthritis Working Group performed a reading exercise from March–April 2016 and presented the results at OMERACT 13 (Whistler, British Columbia, Canada, May 2016). In accordance with the OMERACT handbook[7], no previous calibration tools were found in a literature review by a fellow in this group, which in agreement with the OMERACT executive committee included clinical professionals, methodologists, and healthcare professionals.

## MATERIALS AND METHODS

*Interactive touch-sensitive, Web-based interface*. In OMERACT 12, we demonstrated that a digital image overlay for hip OA BML scoring (by the Hip Inflammation MRI Scoring System) had reliability equivalent to conventional methods and was preferred by readers[8,9]. We developed a Web-based interface suitable for use on touch-sensitive screens, and generated a knee-joint–specific overlay for femur, tibia, and patella. Readers upload or open a sagittal fluid-sensitive [intermediate-weighted fat-saturated (IWFS) or short-tau inversion recovery] knee MRI sequence online (www.carearthritis.com, "Osteoarthritis Imaging," accounts free to registered users). The reader resizes and moves a transparent grid overlay to fit each bone on the central slice where anterior/posterior cruciate ligaments cross, then scrolls through all slices, touching or mouse-clicking each overlay region containing BML. This causes shading to appear (Figure 1), and the Web tool records a score of 1 to indicate that it has been selected. Upon scoring completion, the Web tool outputs a spreadsheet file containing scores 0/1 per region and summary statistics. The maximum possible KIMRISS score across 29 slices includes 763 regions (290 tibia, 377 femur, 96 patella), although no actual knee has BML this extensive. Because regions of templates falling outside bone were not scored, total possible scores will vary slightly for individual patients depending on bone size.

*RETIC tool*. We provided readers with an instructional slide presentation including a scoring atlas giving examples of true BML versus confounders such as volume averaging at condylar edges, subchondral cystic change, and hematopoietic marrow; a video demonstrating KIMRISS scoring (youtu.be/k988FmLVhb0); and the new Web-based RETIC tool. In RETIC training mode, the reader scores cases previously scored by experts. When the reader has finished selecting positive regions, the overlay changes color in each region indicating whether reader and expert scores are concordant/discordant. Intraclass correlation coefficients (ICC) between reader and experts are instantly displayed, allowing real-time calibration and rapid progressive learning with each new case. For RETIC training, 8 cases (2 timepoints each) from the Osteoarthritis Initiative (OAI) data were scored by 2 experienced KIMRISS developers, with discrepancies resolved by consensus.

*Data*. We used publicly available data from the OAI (www.oai.ucsf.edu, v.18). This was a multicenter, prospective observational study of knee OA[10], in which 4796 men and women aged 45–79 years enrolled between 2004 and 2006 underwent annual knee assessment including MRI. We randomly selected 1 knee for each of 20 randomly selected subjects with imaging and clinical information available at baseline and 1-year followup. We scored sagittal IWFS images (repetition/echo times TR/TE 3200/30 ms, matrix 444 × 448, slice thickness 3 mm, field-of-view 159 × 160 mm).

*Reading*. We had 6 readers: 3 musculoskeletal radiologists and 3 rheumatologists (7–30+ yrs of experience). Only 1 reader had previously scored the KIMRISS. Each reader scored BML at tibia, femur, and patella for each selected knee at both timepoints once. Scans from the 2 timepoints were presented to readers together, but readers were blinded to order.

*Statistical analysis*. Interobserver ICC were calculated between BML scores for each reader pair, for BML status, and for change between timepoints. We computed BML scores for the whole joint, for femoral, tibial, and patellar regions, and for subregions intended to match MOAKS regions as closely as possible[2]. We computed the smallest detectable change (SDC) as 1.96 × SEM (standard error of mean change) ÷ $\sqrt{2}$[11].

## RESULTS

The OAI data used for scoring the KIMRISS showed only a light burden of BML, with baseline total scores averaging 21.4/763, or just 2.8% of the maximum possible. The changes after 1 year were small, with change greater than the calculated SDC evident in 54%, 26%, and 29% of the knees for femur, tibia, and patella, respectively (Table 1).

Despite this, observers scored BML with high reliability. ICC for BML status/change averaged 0.84/0.66 for femur and 0.78/0.88 for tibia (Table 1). Patellar BML was poorly assessed (overall status/change ICC 0.42/0.26), but on closer review, 4 readers performed consistently (BML baseline ICC between readers C, D, E, and F = 0.83–0.96) whereas 2 others, a rheumatologist and a radiologist, were outliers (BML baseline ICC between readers A, B, and others = 0.14–0.52). Subregion analysis in MOAKS-like regions was difficult to interpret, with low BML scores averaging < 1 in most subregions, associated with fair to poor ICC. Notable exceptions were tibial intercondylar notch (mean change 3.0, change ICC 0.93) and lateral central tibia (mean 2.7, baseline/change ICC 0.86/0.66). Scoring appeared highly reliable (SDC = 0.7% of maximum score; Table 1).

In the postexercise survey, reading times were 3–12 min

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.



Downloaded on November 11, 2026 from www.jrheum.org

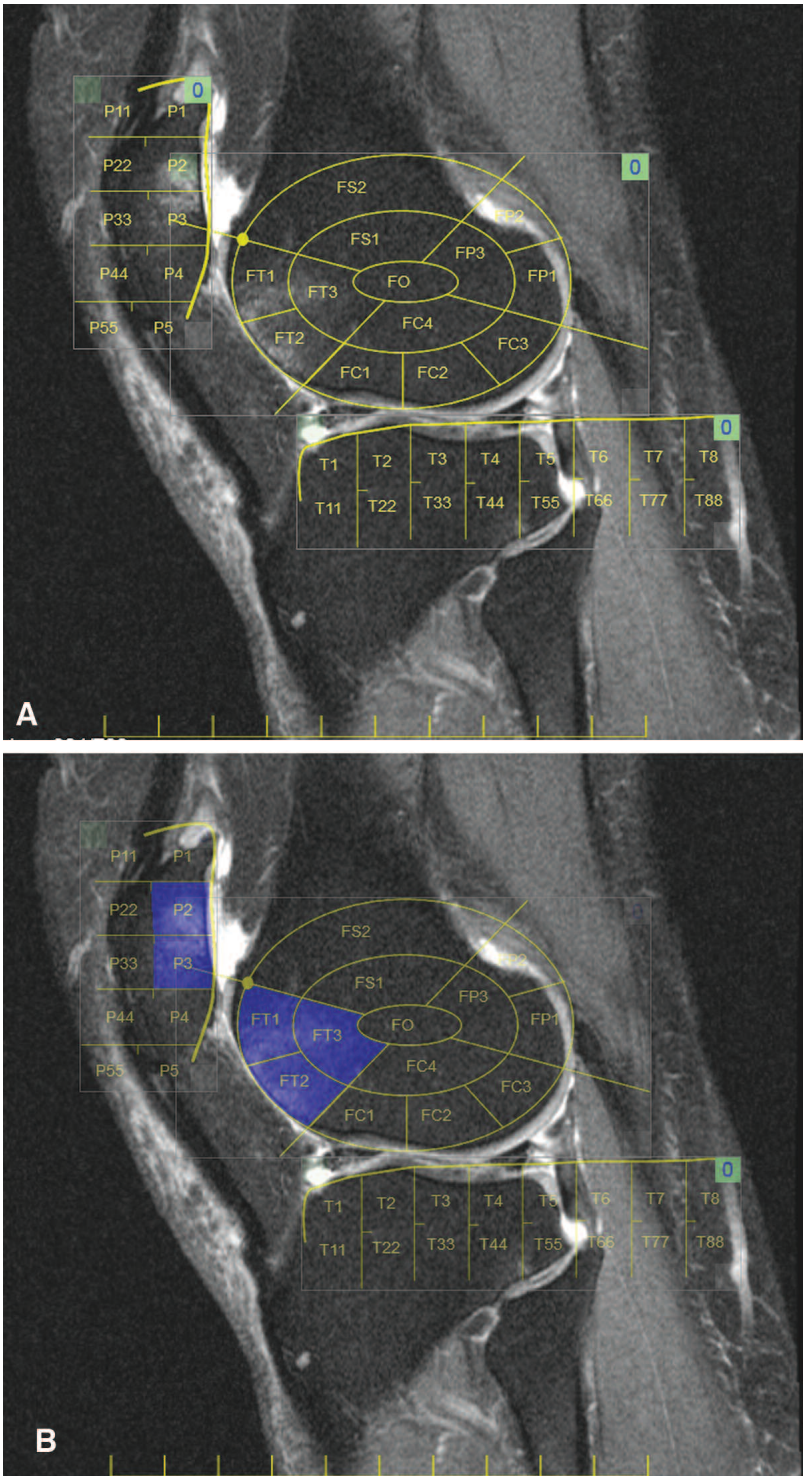


*Figure 1*. Screen captures from KIMRISS Web scoring environment. (A) Overlays in place on case 14 from the current study. On this slice, bone marrow lesions are present in regions P2, P3, FT1, FT2, and FT3. Regions of the template falling outside the bone of interest because of patient size variation are not scored. (B) Screen after the user has touched or mouse-clicked each of these regions. The Web-based scoring system automatically stores each data point (total score of 5 for this slice). KIMRISS: Knee Inflammation Magnetic Resonance Imaging Scoring System; T: tibia; P: patella; FT: femoral trochlea; FS: femoral superior; FP: femoral posterior; FC: femoral central.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.



Downloaded on November 11, 2026 from www.jrheum.org

*Table 1*. Reliability and sensitivity to change for KIMRISS in a multireader exercise scoring 20 knees.

| BML Site | Pooled ICC, All Readers | 95% CI for Pooled ICC | Range of ICC for Reader Pairs | Pooled Mean Score | Pooled SD | Score Range, Individual Readers | SDC | Average % of Knees with Change > SDC |
|---|---|---|---|---|---|---|---|---|
| Femur | | | | | | | | |
| Status | 0.84 | 0.70–0.93 | 0.66–0.97 | 6.6 | 9 | 0–30 | | |
| Change | 0.66 | 0.48–0.84 | 0.36–0.82 | 1.6 | 4.9 | –7 to 10 | 1.5 | 54 |
| Tibia | | | | | | | | |
| Status | 0.78 | 0.64–0.90 | 0.54–0.94 | 6.9 | 11.2 | 0–50 | | |
| Change | 0.88 | 0.78–0.95 | 0.70–0.97 | 6.3 | 18.2 | –8 to 74 | 5.6 | 26 |
| Patella | | | | | | | | |
| Status* | 0.42* | 0.34–0.66 | 0.14–0.99 | 7.9 | 10.1 | 0–31 | | |
| Change | 0.26 | 0.10–0.52 | 0.04–0.98 | –1.5 | 6.6 | –25 to 7 | 2.1 | 29 |
| Total | | | | | | | | |
| Status | 0.62 | 0.41–0.81 | 0.29–0.95 | 21.3 | 17 | 0–54 | | |
| Change | 0.76 | 0.60–0.89 | 0.51–0.95 | 6.4 | 20.5 | –17 to 71 | 6.3 | 41 |

* ICC between readers C, D, E, F: 0.83–0.96; ICC between readers A, B and these 4: 0.14–0.52. KIMRISS: Knee Inflammation Magnetic Resonance Imaging Scoring System; BML: bone marrow lesion; ICC: intraclass correlation coefficients; SDC: smallest detectable change.

and 67% of readers reported that the system was "very user friendly," while none said it was not user friendly.

## DISCUSSION

Our exercise demonstrated the feasibility and high reliability of MRI-defined BML scoring in knee OA by an international group of readers new to the scoring system, self-trained using a standardized interactive feedback calibration module, a Web-based interface with transparent overlay, and touch/click-sensitive screen. Reliability was surprisingly high given the limited BML in the knees studied. The wide scoring range in the OMERACT KIMRISS allowed analysis as a continuous outcome. The KIMRISS was sufficiently reliable to detect small changes < 1% of maximum possible score. Because the KIMRISS is not limited to a periarticular bone, it can also allow the future study of the clinical relevance of nonsubchondral BML or other non-OA disease conditions such as rheumatoid arthritis or avascular necrosis. At OMERACT 13, the MRI in the Arthritis Working Group expressed interest in further validation studies using this methodology.

Because cases were randomly sampled from the OAI observational study, there was a relatively low burden of BML and only minor changes over 1 year. A more comprehensive test of the KIMRISS reliability and sensitivity to change would include the full spectrum of OA severity: cases with large amounts of BML or with substantial changes in BML over time, such as in a therapeutic trial.

We noted specific concerns in scoring the patella, where 4 readers were reliable and 2 readers were outliers. The patella is small and subcutaneous; possible explanations include volume averaging effects and "coil burn" artifact from field inhomogeneity, which can mimic BML. Adjustments to the instructional slide show and changes to the selection of sample cases for RETIC training may reduce these concerns.

Overall, the use of a Web-based scoring interface with RETIC interactive calibration was highly feasible for knee BML scoring and showed high reliability even in new users. Further testing is required in datasets more closely representative of the therapeutic trial setting in which the KIMRISS may be tested for responsiveness and discrimination.

## ACKNOWLEDGMENT

Thanks to Stephanie Belton for her statistical expertise and long hours of analysis, and Joel Paschke for his computer programming skills, which were essential to the development of the Web interface.

## REFERENCES

1. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206-11.
2. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19:990-1002.
3. Lynch JA, Roemer FW, Nevitt MC, Felson DT, Niu J, Eaton CB, et al. Comparison of BLOKS and WORMS scoring systems part I. Cross sectional comparison of methods to assess cartilage morphology, meniscal damage and bone marrow lesions on knee MRI: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2010;18:1393-401.
4. Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SM. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. Arthritis Res Ther 2015;17:152.
5. Bellamy N. Clinimetric concepts in outcome assessment: the OMERACT filter. J Rheumatol 1999;26:948-50.
6. Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingham CO 3rd, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.



Downloaded on November 11, 2026 from www.jrheum.org

J Rheumatol 2014;41:1025-30.
7. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO 3rd, Conaghan P, et al. The OMERACT Handbook. OMERACT 2016. [Internet. Accessed February 15, 2017.] Available from: www.omeract.org/pdf/OMERACT_Handbook.pdf
8. Jaremko JL, Pitts M, Maksymowych WP, Lambert RG. Development of image overlay and knowledge transfer module technologies aimed at enhancing feasibility and external validation of magnetic resonance imaging-based scoring systems. J Rheumatol 2016;43:223-31.
9. Maksymowych WP, Pitts M, Budak MJ, Gracey D, Lambert RG, McDougall D, et al. Development and preliminary validation of a digital overlay-based learning module for semiquantitative evaluation of magnetic resonance imaging lesions in osteoarthritis of the hip. J Rheumatol 2016;43:232-8.
10. Eckstein F, Wirth W, Nevitt MC. Recent advances in osteoarthritis imaging—the osteoarthritis initiative. Nat Rev Rheumatol 2012;8:622-30.
11. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2017. All rights reserved.



Downloaded on November 11, 2026 from www.jrheum.org