

RMD
OpenRheumatic &
Musculoskeletal
Diseases

ORIGINAL ARTICLE

Preliminary validation of the Knee Inflammation MRI Scoring System (KIMRISS) for grading bone marrow lesions in osteoarthritis of the knee: data from the Osteoarthritis Initiative

Jacob L Jaremko,¹ Dean Jeffery,¹ M Buller,¹ Stephanie Wichuk,² Dave McDougall,¹ Robert GW Lambert,¹ Walter P Maksymowych²

To cite: Jaremko JL, Jeffery D, Buller M, *et al*. Preliminary validation of the Knee Inflammation MRI Scoring System (KIMRISS) for grading bone marrow lesions in osteoarthritis of the knee: data from the Osteoarthritis Initiative. *RMD Open* 2017;**3**:e000355. doi:10.1136/rmdopen-2016-000355

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/rmdopen-2016-000355>).

Received 27 August 2016
Revised 8 December 2016
Accepted 23 December 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Jacob L Jaremko;
jjaremko@ualberta.ca

ABSTRACT

Objective: Bone marrow lesions (BML) are an MRI feature of osteoarthritis (OA) offering a potential target for therapy. We developed the Knee Inflammation MRI Scoring System (KIMRISS) to semiquantitatively score BML with high sensitivity to small changes, and compared feasibility, reliability and responsiveness versus the established MRI Osteoarthritis Knee Score (MOAKS).

Methods: KIMRISS incorporates a web-based graphic overlay to facilitate detailed regional BML scoring. Observers scored BML by MOAKS and KIMRISS on sagittal fluid-sensitive sequences. Exercise 1 focused on interobserver reliability in Osteoarthritis Initiative observational data, with 4 readers (two experienced/two new to KIMRISS) scoring BML in 80 patients (baseline/1 year). Exercise 2 focused on responsiveness in an open-label trial of adalimumab, with 2 experienced readers scoring BML in 16 patients (baseline/12 weeks).

Results: Scoring time was similar for KIMRISS and MOAKS. Interobserver reliability of KIMRISS was equivalent to MOAKS for BML status (ICC=0.84 vs 0.79), but consistently better than MOAKS for change in BML: Exercise 1 (ICC 0.82 vs 0.53), Exercise 2 (ICC 0.90 vs 0.32), and in new readers (0.87–0.92 vs 0.32–0.51). KIMRISS BML was more responsive than MOAKS BML: post-treatment BML improvement in Exercise 2 reached statistical significance for KIMRISS (SRM –0.69, $p=0.015$), but not MOAKS (SRM –0.12, $p=0.625$). KIMRISS BML also more strongly correlated to WOMAC scores than MOAKS BML ($r=0.80$ vs 0.58 , $p<0.05$).

Conclusions: KIMRISS BML scoring was highly feasible, and was more reliable for assessment of change and more responsive to change than MOAKS BML for expert and new readers.

INTRODUCTION

As new osteoarthritis (OA) treatment options emerge focused on inflammation,¹ there is

Key messages

What is already known about this subject?

- The extent of knee osteoarthritis can be objectively characterised by semiquantitative grading of MRI features, to help understand which patients will have progressive disease and which treatments are effective. Bone marrow lesion (BML) is an important measure of active disease, but existing scoring systems have limited sensitivity to small changes over time.

What does this study add?

- This study introduces Knee Inflammation MRI Scoring System (KIMRISS), a scoring system that uses an electronic overlay to allow readers to record rapid touch-based or click-based binary scoring decisions for many small regions of bone, in a convenient web-based environment. This precise and detailed scoring would be impractical by traditional manual scoring methods. We show that KIMRISS is feasible, has reliability equivalent or higher than the current MRI Osteoarthritis Knee Score scoring system, and has higher sensitivity to interval changes. The novel combination of electronic overlays and direct on-screen scoring via web-based interface can also be applied in future to other types of image-based scoring, in other body parts and other disease processes. It may be an important tool for external knowledge transfer of newly developed scoring platforms based on imaging.

How might this impact on clinical practice?

- Use of the highly sensitive KIMRISS scoring system could in future allow osteoarthritis clinical trials to be performed more cost-effectively with fewer patients needed to achieve statistically significant results. Web-based scoring with digital overlays could make semiquantitative scoring faster and easier to teach to new readers, facilitating clinical trials. The detailed subregional scoring data provided by KIMRISS may allow new insights into which distributions of BML predict a high risk for rapid progression of osteoarthritis.

an increasing need for accurate and reproducible scoring methods to measure disease severity and treatment response. Bone marrow lesions (BML) seen on MRI may offer a target for therapy. BML are associated with increased risk for subsequent cartilage damage, especially when new or increasing,^{2–6} and in many studies, but not all, BML are associated with pain.^{6–16} Evidence is conflicting as to whether pain severity is correlated to BML size^{6 12 13 15 17 18} or not,^{9 11 19–21} or to BML location.^{12–14} This inconsistency may relate to limitations of the MRI-based BML scoring systems currently in use, particularly regarding relative insensitivity to small lesions or subtle changes.

There are a variety of semiquantitative knee OA scoring systems, with the most commonly used including the Whole-Organ MRI Score (WORMS),²² Boston-Leeds Osteoarthritis Knee Score (BLOKS),²³ and MRI Osteoarthritis Knee Score (MOAKS).²⁴ Other variants exist.^{25–28} BLOKS is a modification of WORMS with increased emphasis on BML appearance.^{29 30} MOAKS combines features of WORMS and BLOKS, with increased subdivision of regional assessment and further changes to BML scoring.²⁴ Additional refinement of semiquantitative BML scoring may be helpful, particularly for longitudinal observational studies or clinical trials in which interval changes are graded to determine which subsets of patients progress or whether a treatment is effective. To increase the sensitivity to small changes in longitudinal data sets, ‘within-grade scoring’ for cartilage and BML has been tested, in which the reader notes whether there has been perceptible interval change that is not sufficient to alter the formal score by a full grade.³¹ The subjective assessment needed for this within-grade scoring has potential to increase inter-reader variability. An alternative to this modification would be to use a more granular scoring system, designed from the outset to be highly sensitive, which could more conveniently and reproducibly record these small changes.

Quantitative knee OA scoring methods, previously focused mainly on cartilage, can also determine BML volume by multiplying BML widths in three dimensions,³² manually drawing contours around BML,³³ or determining regions from initial user inputs using specialised software.³⁴ These methods can reduce observer dependence, but may require dedicated research-protocol MRI sequences, time-consuming manual measurements and/or specific postprocessing software. Semiquantitative scoring is more accessible, but can be complex for readers. Studies directly comparing semiquantitative versus quantitative BML scoring yield conflicting data regarding sensitivity to change.^{35 36}

We have developed the Knee Inflammation MRI Scoring System (KIMRISS) to focus on factors thought to relate most to active disease, considered by many to have an inflammatory component: BML and synovitis-effusion.³⁷ At the hip, BML measured by a similar scoring system, HIMRISS, correlated significantly

to hip pain in early OA.³⁸ KIMRISS and HIMRISS use a novel web-based image overlay for precise region definition. In this study, we compare inter-reader reliability of KIMRISS versus MOAKS BML in observational data (Exercise 1) and responsiveness to change in a therapeutic-trial setting (Exercise 2). We also assess validity by determining associations between BML, pain and function, and progression to arthroplasty.

METHODS

Scoring systems

KIMRISS BML scoring uses a HTML5 web-based interface (free for registered users, <http://www.carearthritis.com>), superimposing an interactive overlay (see online supplementary figures S1 and S2) on a sagittal fluid-sensitive MRI sequence (short- τ inversion recovery or fat-saturated proton-density-weighted). We prefer use of true sagittal sequences but oblique sagittal sequences planned with respect to the anterior cruciate ligament are also acceptable for scoring. For reading exercises in a clinical trial, the reader will open an MRI already uploaded in a sequence prepared by the trial designers. Alternatively, the reader can manually upload an MRI to read directly. The overlay is moved by the reader to fit bone at three slices for the femur and tibia (central slice, medial compartment, lateral compartment). The overlay position is then automatically adjusted by interpolation to best fit other image slices, minimising reading time and user variability. In future this overlay positioning could be fully automated by use of an image segmentation routine. The overlay separates subarticular bone into $\sim 1 \times 1$ cm regions. For 3 mm slice thickness, all knees in pilot data were captured within 29 sagittal slices, giving up to 763 regions (290 tibia, 377 femur, 96 patella). Each region is scored 0 by default. On each slice, the reader touches or mouse-clicks each BML-containing region, updating scores for those regions to 1 and causing them to change colour onscreen for feedback. The resulting BML scores are exportable in comma separated values (CSV) format for analysis. KIMRISS BML scoring is detailed in online supplementary materials.

MOAKS BML is scored in 15 regions (2 patellar, 6 femoral, 7 tibial), based on BML size (none=0, <33% of region=1; 33–66% of region=2, >67%=3), for maximum score $15 \times 3 = 45$ per knee.²⁴ MOAKS also scores the percentage of BML that is non-cystic; since any purely cystic region is scored 0 in KIMRISS we did not evaluate these regions specifically. MOAKS also scores the number of BML in a region; since BML adjacent to each other may appear to merge or separate depending on technical factors and observer perception, this is a difficult parameter to score reproducibly or analyse meaningfully and was not considered for the current study.

Data available

Exercise 1: We used publicly available data (release 18) from the Osteoarthritis Initiative (OAI), a multicentre

observational study of 4796 patients with, or at risk for, OA.³⁹ Our goal was to test scoring system reliability across the full spectrum of knees from no BML to extensive BML. To capture highly symptomatic knees likely to have large amounts of BML, we selected the first 40 consecutive OAI patients who went on to knee steroid injection within 1 year after enrolment. To complete the spectrum we included a non-injection cohort, 40 consecutive OAI patients with no knee steroid injection in year 1, matched to the injection-cohort for age, sex, knee side and Kellgren-Lawrence grade of radiographic OA as scored by OAI investigators. Baseline and 1-year follow-up MRIs were scored for each patient.

Exercise 2: Scans were obtained from an open-label pilot study testing treatment efficacy of adalimumab, a biological anti-inflammatory medication for knee OA (<https://www.clinicaltrials.gov/ct2/show/NCT00686439>). Patients received adalimumab 40 mg by subcutaneous injection on alternate weeks. Study design and patient characteristics have been reported previously.⁴⁰ Baseline and 12-week follow-up MRI scans were available for 16 of the 20 study patients (table 1).

Reading exercises

For Exercise 1, focused on interobserver reliability and feasibility, we had four readers score KIMRISS and MOAKS BML in baseline and 1-year MRI. Two expert readers, musculoskeletal radiologists (6 and 11 years of experience) involved in KIMRISS development, scored all 80 available OAI knees. Two readers new to KIMRISS, radiology residents with no previous semiquantitative scoring experience, reviewed a slide presentation describing KIMRISS, three reference cases and the published manuscript describing MOAKS,²⁴ then scored 40 knees randomly selected from the 80 available.

For Exercise 2, focused on responsiveness, the same two expert readers scored all 16 knees for BML by MOAKS and KIMRISS, at baseline and 12-week follow-up.

Readers in both exercises were blinded to time point (baseline vs follow-up) and all clinical and demographic information.

Association of BML to outcomes

Although this work focused primarily on reliability and responsiveness, as a secondary analysis we performed preliminary testing of validity and potential utility of BML scoring by KIMRISS and MOAKS via comparison with clinical status. This was assessed in both exercises by WOMAC (Western Ontario and McMaster Universities), with maximum possible scores: 20 for worst pain, 68 for poorest function.^{41 42}

Statistics

We used SPSS (IBM, V.20). Descriptive statistics were reported as mean±SD. Given the large scoring ranges of both MOAKS and KIMRISS BML scores, we treated each as a quasi-continuous variable for analysis, and for simplicity, considered the whole-joint total BML score for most analyses. For interobserver reliability, we recorded intraclass correlation coefficients (ICC(3,1), two-way mixed single measures, consistency) between each reader pair. We also generated Bland-Altman plots comparing expert reader scores, and computed smallest detectable change (SDC) for KIMRISS and MOAKS BML based on the 95% CI of interobserver variability of change scores.

For responsiveness, in Exercise 2 we computed standardised response means (SRM) for MOAKS and KIMRISS, and performed paired Student's t-tests to assess for statistical significance of observed changes in BML by KIMRISS or MOAKS.

For validity related to pain and disability, we first assessed bivariate Pearson correlations between MRI BML scores (by KIMRISS and MOAKS) and WOMAC pain and function scores at baseline, at follow-up (Exercise 1: 1 year; Exercise 2: 12 weeks) and in terms of interval change. In each Exercise we then performed multivariate linear regression to determine whether

Table 1 Baseline patient characteristics for two reading exercises

Mean (SD) unless specified	Exercise 1 (OAI)			Exercise 2 (Adalimumab) All (n=16)
	Injection (n=40)	Control (n=40)	All (n=80)	
Age (years)	62.3 (8.5)	62.3 (8.6)	62.3 (8.6)	57.8 (10.2)
Males (%)	8 (20%)	8 (20%)	16 (20%)	7 (35%)
Symptom duration (years)	NA	NA	NA	4.7 (4.2)
WOMAC pain (max. 20)	5.0 (3.5)	3.2 (3.9)	4.1 (3.8)	11.9 (3.5)
WOMAC function (max. 68)	24.8 (16.6)	15.6 (17.3)	20.2 (17.5)	43.6 (12.4)
Kellgren-Lawrence Grade	2.9 (1.1)	2.8 (1.1)	2.9 (1.1)	2.8 (0.4)
K-L grade 0: n (%)	1 (2.5%)	1 (2.5%)	2 (2.5%)	0 (0%)
K-L grade 1: n (%)	5 (12.5%)	5 (12.5%)	10 (12.5%)	0 (0%)
K-L grade 2: n (%)	6 (15%)	6 (15%)	12 (15%)	4 (25%)
K-L grade 3: n (%)	16 (40%)	16 (40%)	32 (40%)	12 (75%)
K-L grade 4: n (%)	12 (30%)	12 (30%)	24 (30%)	0 (0%)

Max., maximum; NA, not available; OAI, Osteoarthritis Initiative; WOMAC, Western Ontario and McMaster Universities.

baseline BML or change in BML predicted change in WOMAC pain scores, adjusting for variables correlating significantly in the univariate analysis as well as other potential confounders including age, sex, symptom duration, Kellgren-Lawrence (K-L) grade, and baseline WOMAC pain scores when change was the dependent variable.

RESULTS

Exercise 1

The observed BML scores were much lower than the theoretical maximum scores for KIMRISS and MOAKS, since only portions of even highly symptomatic knees contain BML. KIMRISS and MOAKS BML scores were highly correlated ($r=0.89$). There were non-significant trends towards more BML in the injection group than non-injection group, whether at baseline or 1-year, and whether scored by KIMRISS or MOAKS (table 2). There was also a non-significant trend towards increasing BML scores (ie, worsening disease) in the 1-year follow-up period, by KIMRISS or MOAKS: for example, KIMRISS BML increased in 21/40 non-injected knees and 23/40 injected knees, decreasing in 11 and 13 knees, respectively.

Reading times varied by reader experience, but averaged 3–8 min per scan for KIMRISS and MOAKS (including KIMRISS template sizing and moving, <0.5 min). A knee with no BML could be scored in under 1 min by either method, and scoring times for a knee with mild or moderate BML were similar for the two methods. KIMRISS scoring times were longer than MOAKS for the most severely arthritic knees with very extensive BML, due to the number of clicks required (eg, 10 min vs 7 min).

Reliability for expert users: BML change scores from baseline to follow-up were substantially more reliably generated using KIMRISS than MOAKS (ICC 0.82 vs 0.53). KIMRISS ICC was also slightly higher than MOAKS for

baseline BML status score (0.84 vs 0.79). The SDC was a smaller proportion of the maximum scoring range for KIMRISS than for MOAKS (24.6/763=3.2%, vs 2.3/45=5.1%) (table 2).

Bland-Altman plots comparing expert scores at baseline showed relatively narrow reader differences for KIMRISS in the most common scoring range and wider reader differences across the scoring range for MOAKS. Similar findings were observed for change scores (see online supplementary figure S3). There were small systematic differences between KIMRISS scoring in which one expert appeared to score slightly higher than the other, and some mild heteroscedasticity (change scores differed proportionately more widely in the few cases with large changes than in the majority with small changes).

Reliability for new versus expert readers: Interobserver reliability between scoring by each new reader (N1 and N2) and the average of the two expert reader scores is summarised in table 3. While both new readers assessed BML status scores reliably by either KIMRISS (ICC 0.89–0.98) or MOAKS (ICC 0.87–0.92), change in BML was substantially more reliably assessed by new readers using KIMRISS (ICC 0.87–0.92) than MOAKS (0.32–0.51).

Exercise 2

Likely because OAI patients generally showed little change between baseline and 1-year scans, tests of responsiveness in Exercise 1 showed no statistically significant results. To more meaningfully compare responsiveness of KIMRISS versus MOAKS BML scoring, in Exercise 2 we scored MRI from a small clinical-trial cohort of patients who generally demonstrated substantial improvement in BML 12 weeks following potent anti-inflammatory therapy (figure 1).

Responsiveness: Patients in Exercise 2 showed more active arthropathy at baseline than those in Exercise 1 (KIMRISS BML average 37.3 vs 21.2). Twelve weeks post

Table 2 BML scores in Exercise 1 (40 knees in the injection cohort and 40 matched knees in the control cohort) and Exercise 2 (16 knees)

	BML scores: mean (SD)			SDC (% of max)	p Value	SRM
	Baseline	Follow-up	Change			
EXERCISE 1 (1-year follow-up)						
KIMRISS (injection)	26.7 (21.6)	33.7 (26.4)	7.0 (20.1)	24.6 (3.2%)	0.876*	
KIMRISS (control)	15.6 (17.2)	21.9 (28.6)	6.4 (15.4)			
MOAKS (injection)	4.7 (2.5)	5.0 (2.5)	0.3 (1.7)	2.3 (5.1%)	0.507*	
MOAKS (control)	3.1 (2.3)	3.6 (3.0)	0.5 (1.6)			
EXERCISE 2 (12-week follow-up)						
KIMRISS	37.3 (36.7)	29.8 (34.9)	-7.5 (10.9)	15.1 (1.9%)	0.015†	-0.69
MOAKS	6.3 (0.9)	6.1 (0.8)	-0.2 (0.4)	0.55 (1.2%)	0.625†	-0.12

All readings are the average of two expert reader scores.

Bold=statistically significant.

*For the difference in BML change scores between injection and control groups.

†For the difference between baseline and follow-up BML scores.

BML, bone marrow lesions; SDC, smallest detectable change; SRM, standardised response mean, not calculated for Exercise 1 (not meaningful for observational data).

Table 3 Reliability of expert readers, composed of average of readings by two experts (80 patients Exercise 1; 16 patients Exercise 2)

	ICC (95% CI) for BML Scores			Exercise 2 (Adalimumab) N=16 Between experts
	Exercise 1			
	(OAI) N=80 Between experts	(OAI) N=40 N1 vs expert	N2 vs expert	
KIMRISS				
Status	0.84 (0.66 to 0.91)	0.89 (0.79 to 0.94)	0.98 (0.96 to 0.99)	0.97 (0.90 to 0.99)
Change	0.82 (0.73 to 0.88)	0.87 (0.77 to 0.93)	0.92 (0.85 to 0.96)	0.90 (0.73 to 0.96)
MOAKS				
Status	0.79 (0.69 to 0.86)	0.81 (0.67 to 0.90)	0.87 (0.77 to 0.93)	0.67 (0.25 to 0.87)
Change	0.53 (0.35 to 0.68)	0.32 (0.01 to 0.57)	0.51 (0.23 to 0.70)	0.32 (-0.18 to 0.69)

Reliability: novice readers versus expert readers (40 randomly selected patients from the 80 patient Exercise 1 cohort). BML, bone marrow lesions; ICC, Intraclass Correlation Coefficients; KIMRISS, Knee Inflammation MRI Scoring System; MOAKS, MRI Osteoarthritis Knee Score; OAI, Osteoarthritis Initiative.

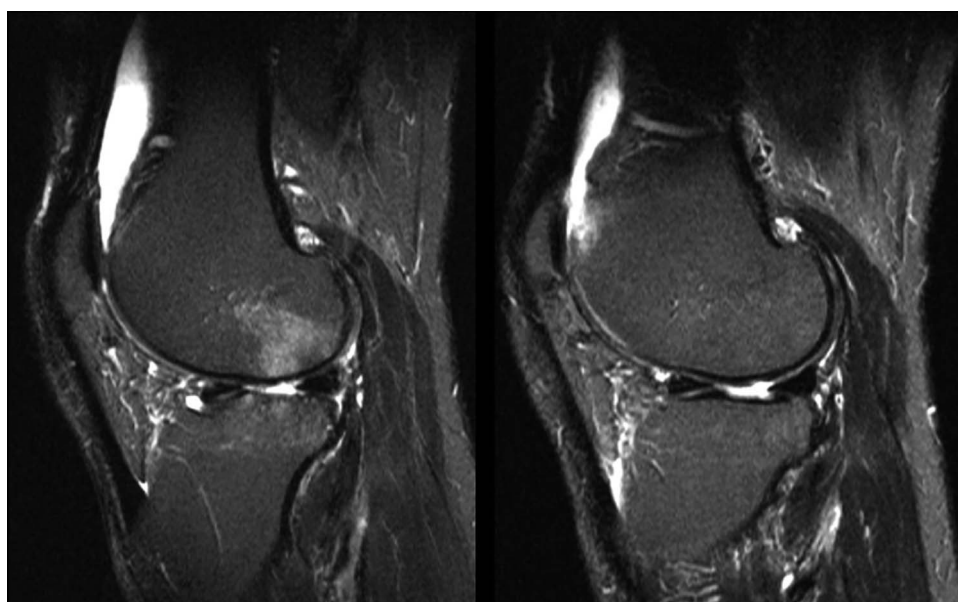


Figure 1 Decrease in BML post-therapy. Sagittal T2 fat-saturated sequences at baseline (left) and 12-week follow-up post Adalimumab treatment (right). Corresponding to the visually obvious improvement in the T2-intense bone marrow lesions in the femur and tibia, in this knee the semiquantitative BML scores dropped from 43 to 21 (KIMRISS) and from 9 to 6 (MOAKS), while WOMAC pain score improved from 16 to 8 and WOMAC function score improved from 64 to 32. BML, bone marrow lesions; KIMRISS, Knee Inflammation MRI Scoring System; MOAKS, MRI Osteoarthritis Knee Score; WOMAC, Western Ontario and McMaster Universities.

anti-inflammatory therapy, KIMRISS demonstrated a statistically significant decrease in BML, ($p=0.015$, $SRM=-0.69$), while the decrease in MOAKS BML score was smaller ($SRM=-0.12$) and not statistically significant (table 2).

Reliability: The changes in BML in Exercise 2 were substantially more reliably assessed by KIMRISS than by MOAKS (ICC 0.90 vs 0.32, table 3). KIMRISS also appeared somewhat more reliable than MOAKS for baseline BML scoring (ICC 0.97 vs 0.67, table 3).

Validity: correlating BML to clinical findings

For OAI data (Exercise 1), no significant bivariate correlations were found between BML scores (KIMRISS,

MOAKS) and WOMAC scores (pain, function). For adalimumab trial data (Exercise 2), baseline WOMAC pain and MRI BML were significantly correlated when measured by KIMRISS at the femur ($r=0.51$, $p=0.048$) or tibia ($r=0.56$, $p=0.023$), but not when measured by a combined whole-joint KIMRISS score ($r=0.48$, $p=0.06$) or when measured by MOAKS ($r=0.36$, $p=0.17$). Changes in BML scores did not significantly correlate to changes in WOMAC scores. At 12-week follow-up, WOMAC scores correlated more strongly to residual BML at the tibia when measured by KIMRISS (pain $r=0.80$, function $r=0.79$, $p<0.0001$) than MOAKS (pain $r=0.58$, $p=0.02$; function $r=0.46$, not significant). Correlations to femoral BML were no longer significant.

DISCUSSION

We designed the new KIMRISS scoring system to focus on potentially reversible MRI biomarkers of active knee OA, and to facilitate scoring by online interface. KIMRISS evaluates BML and effusion/synovitis, which relate more strongly to symptoms than other MRI features such as cartilage changes.²⁴ The OMERACT filter⁴³ evaluates imaging-based scoring systems for truth (validity), discrimination (reliability/responsiveness) and feasibility. Here we compared reliability and responsiveness of scoring BML by KIMRISS versus the existing MOAKS scoring system, in observational and therapeutic-trial settings. (We did not assess the effusion component of KIMRISS or MOAKS scoring here). Our data also allowed us to make some preliminary hypothesis-generating observations regarding validity of BML scoring as related to OA disease status and outcomes, and to comment on scoring feasibility.

Reliability: Interobserver reliability was high and similar for KIMRISS and MOAKS BML scoring for disease status (ICC 0.84 vs 0.79), similar to published inter-rater ICC for MOAKS BML (0.74–1.00).²⁴ However, reliability was consistently substantially better for KIMRISS than MOAKS in the crucial assessment of change in BML, whether for small changes in observational data (Exercise 1, ICC 0.82 vs 0.53), or larger changes in a therapeutic trial (Exercise 2, ICC 0.90 vs 0.32). The greater reliability of KIMRISS was particularly striking when new readers scored change in BML (ICC 0.87–0.92 for KIMRISS vs 0.32–0.51 MOAKS). Particularly for new readers, it seems that the many simple binary scoring decisions (BML yes/no) in KIMRISS are somewhat more reliably made than the fewer but more challenging scoring decisions in MOAKS, which requires readers to estimate the percentage volume involved by BML and cystic change within more anatomically complex regions.

Responsiveness: BML scoring by KIMRISS was more responsive than MOAKS, showing a statistically significant effect of therapy in Exercise 2 that was not demonstrated by MOAKS (BML decrease by KIMRISS SRM -0.69 , $p=0.015$, by MOAKS SRM -0.12 , $p=0.625$). This is due to the greater scoring range of KIMRISS (several hundred small regions) than MOAKS (12 larger regions). Acknowledging the low sensitivity of current scoring systems (BLOKS/WORMS/MOAKS) for small longitudinal changes, others have attempted to compensate by use of ‘within-grade scoring’.³¹ The more granular scoring approach used in KIMRISS offers a more objective alternative that ought to show less inter-reader variability than ‘within-grade scoring’ where readers are asked to rate whether a significant perceptible change has occurred or not. Future comparative studies could test this hypothesis.

Validity: A rationale for studying BML is that it should relate to clinical status and outcomes in OA. Although the non-randomised patient populations in both Exercises were not ideally suited for analysing these

relations, we observed significant correlations, stronger when scored by KIMRISS than MOAKS. For example, Exercise 2 WOMAC scores correlated significantly to KIMRISS BML at baseline ($r=0.51$ – 0.56) and post-treatment ($r=0.8$), with weaker correlations to MOAKS BML (not significant at baseline, at most $r=0.58$ post-treatment). Correlations were higher for KIMRISS scores in femur or tibia alone than when combined. It is possible that BML concentrated in one bone or region may be more clinically meaningful than BML spread through a joint. Concordant with these results, others have also reported significant associations between BML size and knee pain.⁶ In a systematic review in 2011, 63% of studies found significant positive associations between BML and knee pain, with ORs 2.0 to 5.0.⁷ In our study, correlations between BML and pain/function were stronger in Exercise 2 (therapeutic trial) than Exercise 1 (observational data). This may be because BML is known to fluctuate over weeks in individuals being observed with OA.⁴⁴ Given the mix of physiological and treatment effects on BML, we agree with others that BML may be a more appropriate outcome measure in clinical trials than observational studies.⁴⁴

The KIMRISS method includes scoring of non-articular bone (eg, regions F0, FS1, FS2), in which BML may be less relevant to the OA disease process than in subchondral bone. Although this may initially seem counter-intuitive, we have intentionally included these regions in scoring, for two reasons: it allows future analysis comparing BML in subchondral and non-subchondral locations which can clarify whether non-subchondral BML in fact has any clinical relevance in OA, and it increases the potential applicability of the KIMRISS system to other disease processes such as inflammatory arthropathies and avascular necrosis, in which non-subchondral BML may be highly clinically meaningful.

KIMRISS scores hundreds of tiny subregions, which can be combined in a variety of ways, each with advantages and disadvantages. For this initial study, focused on feasibility and reliability, we have simply summed all subregions in each bone at the knee. For future studies addressing clinical questions in larger patient populations, such a whole-bone or whole-joint total score may not be optimal since changes in some regions may balance out others; changes in some regions such as non-subchondral bone may not be as clinically relevant as in others; or the scoring system may be over-sensitive to tiny changes which are not clinically meaningful. In future studies in large patient populations, different approaches to combining KIMRISS scores can be tested. For example, perhaps a score including only the medial and lateral weight-bearing subchondral regions would correlate best to clinical symptoms, and perhaps a minimum threshold of change likely to be clinically relevant can be established. KIMRISS should be thought of mainly as a reliable means of acquiring the raw data regarding extent and location of knee BML, which can

then be combined in ways optimised to each clinical question.

Correlations between BML and change in pain were not significant in this study, but in Exercise 2, in which this was most meaningfully assessed, the sample size was small. Given the strength of correlations between BML and pain status scores at baseline and follow-up, a larger study might also show significant correlations between change scores. This significant association between BML and clinical status in a therapeutic-trial setting reconfirms the clinical relevance of BML, supporting its use in MRI-based OA scoring and validating the concept behind KIMRISS.

Feasibility: KIMRISS scoring would have been highly cumbersome using a pencil or spreadsheet, but the web-based environment using mouse or touch-screen input on a digital overlay allowed KIMRISS to be rapid and user-friendly, with intuitive binary scoring. KIMRISS and MOAKS had similar scoring times of 3–8 min per scan, perhaps because time spent clicking more small subregions in KIMRISS balanced the time spent mentally calculating the percentage of each region involved in BML for MOAKS, at least for mild to moderate OA. Scoring of severely arthritic knees with extensive BML was slower using KIMRISS than MOAKS due to the many clicks/touches needed.

Feasibility of KIMRISS semiquantitative BML scoring may be equivalent or preferable to fully quantitative methods, which can involve a similar level of reader supervision and expertise. In automated quantitation, the reader may need to place seed points within BML and/or adjust signal intensity thresholds to characterise BML accurately; the reader doing this may as well mouse-click or touch regions of BML on the KIMRISS digital overlay. Also, semiquantitative scoring such as by KIMRISS does not require specialised high-resolution MRI sequences or advanced workstation postprocessing.

The KIMRISS BML system we have described does not assess other potential MRI measures of treatment response considered in other scoring systems; depending on goals of scoring, these could be added. For example, in some future OA studies it may be useful to perform MOAKS grading of most periarticular pathology, while substituting the more detailed KIMRISS BML scoring for the MOAKS BML component. Grading of effusion/synovitis is also part of KIMRISS, but limiting assessment in this study to BML maximised ease of use while maintaining responsiveness.

This study had limitations. The sample size was small, particularly in Exercise 2. Although OAI data was collected by strict protocols, it was not collected by us and information such as exactly when steroid injection occurred in each knee was unavailable. Our non-randomised inclusion criteria in Exercise 1 (injection/non-injection cohorts) limit generalisability of inferences relating patient characteristics and long-term outcomes. However, this was not our purpose; we designed this

study to efficiently compare two BML scoring systems across the widest possible spectrum of OA severity. Further studies in which KIMRISS scoring is applied to larger, randomised data sets can more fully assess associations between BML, other features of OA, and outcomes. We noted minor systematic interobserver variability between expert KIMRISS readings. This may be due to varying thresholds between readers for recording positive BML, and should be correctable with more standardised user training in future studies. As with any binary scoring system, there is a risk of over-estimation or under-estimation of change, for example, if BML only partially fills a KIMRISS region at baseline but fills more of this region at follow-up the score (ie, 1) would not change in this region. However, given the much smaller region sizes in KIMRISS than in MOAKS, the impact of such estimation errors should be less in KIMRISS. Finally, it is acknowledged that multiple small BML may have a different clinical significance than a few large BML, which is not reflected in the analysis methodology of this study where we assessed primarily the whole-joint total BML score and the aggregated subscores in 12 large regions corresponding to those used in MOAKS. Larger sample sizes would allow more detailed sub-regional analysis, so that BML distribution and changes could be defined in a more refined manner than in this pilot study.

In summary, scoring of BML by KIMRISS showed advantages over MOAKS, more reliably assessing change in BML, particularly for novices, showing greater responsiveness, demonstrating a significant effect of therapy in a small cohort and correlating more strongly to clinical outcomes. These results support the use of KIMRISS BML scoring as an imaging outcome measure in OA clinical research, and suggest that its use could aid in our understanding of OA progression.

Author affiliations

¹Department of Radiology & Diagnostic Imaging, University of Alberta Hospital, Edmonton, Alberta, Canada

²Faculty of Medicine, Division of Rheumatology, University of Alberta Hospital, Edmonton, Alberta, Canada

Contributors J LJ, RGWL and WPM were involved in design and conception of the study. DJ, MB, SW and DM contributed by acquiring and analysing the data. J LJ, DJ and MB contributed by drafting the manuscript. All authors made substantial contribution by reviewing the manuscript, approving the final version and agreeing to be accountable for the work.

Funding Costs of the portion of the study involving patients who received adalimumab were supported by an investigator-initiated study supported by AbbVie. AbbVie played no role in the design of data acquisition, analysis or interpretation of the data or the development of related manuscripts including this manuscript. J LJ has unrestricted support from the Capital Health Endowment in Diagnostic Imaging.

Competing interests None declared.

Ethics approval University of Alberta Health Research Ethics Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Jasin HE. Immune mechanisms in osteoarthritis. *Semin Arthritis Rheum* 1989;18:86–90.
- Hunter DJ, Zhang Y, Niu J, *et al.* Increase in bone marrow lesions associated with cartilage loss: a longitudinal magnetic resonance imaging study of knee osteoarthritis. *Arthritis Rheum* 2006;54:1529–35.
- Raynauld JP, Martel-Pelletier J, Berthiaume MJ, *et al.* Correlation between bone lesion changes and cartilage volume loss in patients with osteoarthritis of the knee as assessed by quantitative magnetic resonance imaging over a 24-month period. *Ann Rheum Dis* 2008;67:683–8.
- Kothari A, Guermazi A, Chmiel JS, *et al.* Within-subregion relationship between bone marrow lesions and subsequent cartilage loss in knee osteoarthritis. *Arthritis Care Res (Hoboken)* 2010;62:198–203.
- Davies-Tuck ML, Wluka AE, Forbes A, *et al.* Development of bone marrow lesions is associated with adverse effects on knee cartilage while resolution is associated with improvement—a potential target for prevention of knee osteoarthritis: a longitudinal study. *Arthritis Res Ther* 2010;12:R10.
- Driban JB, Price L, Lo GH, *et al.* Evaluation of bone marrow lesion volume as a knee osteoarthritis biomarker—longitudinal relationships with pain and structural changes: data from the Osteoarthritis Initiative. *Arthritis Res Ther* 2013;15:R112.
- Yusuf E, Kortekaas MC, Watt I, *et al.* Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. *Ann Rheum Dis* 2011;70:60–7.
- Lo GH, McAlindon TE, Niu J, *et al.* Bone marrow lesions and joint effusion are strongly and independently associated with weight-bearing pain in knee osteoarthritis: data from the osteoarthritis initiative. *Osteoarthritis Cartilage* 2009;17:1562–9.
- Felson DT, Chaisson CE, Hill CL, *et al.* The association of bone marrow lesions with pain in knee osteoarthritis. *Ann Intern Med* 2001;134:541–9.
- Baum T, Joseph GB, Arulanandan A, *et al.* Association of magnetic resonance imaging-based knee cartilage T2 measurements and focal knee lesions with knee pain: data from the Osteoarthritis Initiative. *Arthritis Care Res (Hoboken)* 2012;64:248–55.
- Buck FM, Hoffmann A, Hofer B, *et al.* Chronic medial knee pain without history of prior trauma: correlation of pain at rest and during exercise using bone scintigraphy and MR imaging. *Skeletal Radiol* 2009;38:339–47.
- Foong YC, Khan HI, Blizzard L, *et al.* The clinical significance, natural history and predictors of bone marrow lesion change over eight years. *Arthritis Res Ther* 2014;16:R149.
- Kim IJ, Kim DH, Jung JY, *et al.* Association between bone marrow lesions detected by magnetic resonance imaging and knee pain in community residents in Korea. *Osteoarthritis Cartilage* 2013;21:1207–13.
- Stefanik JJ, Gross KD, Guermazi A, *et al.* The relation of MRI-detected structural damage in the medial and lateral patellofemoral joint to knee pain: the Multicenter and Framingham Osteoarthritis Studies. *Osteoarthritis Cartilage* 2015;23:565–70.
- Torres L, Dunlop DD, Peterfy C, *et al.* The relationship between specific tissue lesions and pain severity in persons with knee osteoarthritis. *Osteoarthritis Cartilage* 2006;14:1033–40.
- Felson DT, Niu J, Guermazi A, *et al.* Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. *Arthritis Rheum* 2007;56:2986–92.
- Dore D, Quinn S, Ding C, *et al.* Natural history and clinical significance of MRI-detected bone marrow lesions at the knee: a prospective study in community dwelling older adults. *Arthritis Res Ther* 2010;12:R223.
- Zhang Y, Nevitt M, Niu J, *et al.* Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. *Arthritis Rheum* 2011;63:691–9.
- Ai F, Yu C, Zhang W, *et al.* MR imaging of knee osteoarthritis and correlation of findings with reported patient pain. *J Huazhong Univ Sci Technolog Med Sci* 2010;30:248–54.
- Link TM, Steinbach LS, Ghosh S, *et al.* Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. *Radiology* 2003;226:373–81.
- Wildi LM, Raynauld JP, Martel-Pelletier J, *et al.* Relationship between bone marrow lesions, cartilage loss and pain in knee osteoarthritis: results from a randomised controlled clinical trial using MRI. *Ann Rheum Dis* 2010;69:2118–24.
- Peterfy CG, Guermazi A, Zaim S, *et al.* Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage* 2004;12:177–90.
- Hunter DJ, Lo GH, Gale D, *et al.* 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.
- Hunter DJ, Guermazi A, Lo GH, *et al.* Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* 2011;19:990–1002.
- Kornat PR, Ceulemans RY, Kroon HM, *et al.* MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiol* 2005;34:95–102.
- Sowers M, Karvonen-Gutierrez CA, Jacobson JA, *et al.* Associations of anatomical measures from MRI with radiographically defined knee osteoarthritis score, pain, and physical functioning. *J Bone Joint Surg Am* 2011;93:241–51.
- Roemer FW, Frobell R, Lohmander LS, *et al.* Anterior Cruciate Ligament OsteoArthritis Score (ACLOAS): longitudinal MRI-based whole joint assessment of anterior cruciate ligament injury. *Osteoarthritis Cartilage* 2014;22:668–82.
- Roemer FW, Guermazi A, Trattnig S, *et al.* Whole joint MRI assessment of surgical cartilage repair of the knee: cartilage repair osteoarthritis knee score (CROAKS). *Osteoarthritis Cartilage* 2014;22:779–99.
- Lynch JA, Roemer FW, Nevitt MC, *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.
- Felson DT, Lynch J, Guermazi A, *et al.* Comparison of BLOKS and WORMS scoring systems part II. Longitudinal assessment of knee MRIs for osteoarthritis and suggested approach based on their performance: data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 2010;18:1402–7.
- Roemer FW, Nevitt MC, Felson DT, *et al.* Predictive validity of within-grade scoring of longitudinal changes of MRI-based cartilage morphology and bone marrow lesion assessment in the tibio-femoral joint—the MOST study. *Osteoarthritis Cartilage* 2012;20:1391–8.
- Driban JB, Lo GH, Lee JY, *et al.* Quantitative bone marrow lesion size in osteoarthritic knees correlates with cartilage damage and predicts longitudinal cartilage loss. *BMC Musculoskelet Disord* 2011;12:217.
- Ratzlaff C, Guermazi A, Collins J, *et al.* A rapid, novel method of volumetric assessment of MRI-detected subchondral bone marrow lesions in knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:806–14.
- Pang J, Driban JB, Destenaves G, *et al.* Quantification of bone marrow lesion volume and volume change using semi-automated segmentation: data from the osteoarthritis initiative. *BMC Musculoskelet Disord* 2013;14:3.
- Eckstein F, Guermazi A, Roemer FW. Quantitative MR imaging of cartilage and trabecular bone in osteoarthritis. *Radiol Clin North Am* 2009;47:655–73.
- Stahl R, Jain SK, Lutz J, *et al.* Osteoarthritis of the knee at 3.0 T: comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study. *Skeletal Radiol* 2011;40:1315–27.
- Hunter DJ, Zhang W, Conaghan PG, *et al.* Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. *Osteoarthritis Cartilage* 2011;19:557–88.
- Maksymowych WP, Cibere J, Loeuille D, *et al.* Preliminary validation of 2 magnetic resonance image scoring systems for osteoarthritis of the hip according to the OMERACT filter. *J Rheumatol* 2014;41:370–8.
- Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis Cartilage* 2008;16:1433–41.

40. Maksymowych WP, Russell AS, Chiu P, *et al.* Targeting tumour necrosis factor alleviates signs and symptoms of inflammatory osteoarthritis of the knee. *Arthritis Res Ther* 2012;14:R206.
41. Bellamy N, Buchanan WW, Goldsmith CH, *et al.* Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
42. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. *J Rheumatol* 2002;29:2473–6.
43. Boers M, Brooks P, Strand CV, *et al.* The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198–9.
44. Felson DT, Parkes MJ, Marjanovic EJ, *et al.* Bone marrow lesions in knee osteoarthritis change in 6–12 weeks. *Osteoarthritis Cartilage* 2012;20:1514–18.