A Rapid, Novel Method of Volumetric Assessment of MRI-Detected Subchondral Bone Marrow Lesions in Knee Osteoarthritis

C. Ratzlaff[1], **A. Guermazi**[2], **J. Collins**[1], **J.N. Katz**[1], **E. Losina**[1], **C. Vanwyngaarden**[3], **R. Russell**[1], **T. Iranpour**[1], and **J. Duryea**[1]

[1]Brigham and Women's Hosp., Boston, MA

[2]Boston Univ. Sch. of Med., Boston, MA

[3]Northern Lights Regional Health Care Centre, Alberta, Canada

Abstract

Purpose—To assess reliability and validity of a semi-automated quantitative method for osteoarthritis (OA)-related bone marrow lesion (BML) assessment in the femur and tibia.

Methods—In a cross-sectional study of subjects with knee OA, we examined concurrent criterion and clinical validation of a novel method of semi-automated quantitative BML measurement. The primary outcome was total segmented BML volume in femoral and tibial medial and lateral knee compartments. Criterion validation was examined through comparison of BML volumes with Whole-Organ Magnetic Resonance Imaging Score (WORMS) scoring. Clinical validation was examined via associations of tibial and femoral BML volume with the Western Ontario and McMaster University OA Index weight-bearing pain questions.

Results—Among the 115 subjects, mean age was 62 years, mean BMI 30.4 (kg/m^2), 84% were white and 52% male. The ICC for intra-reader reliability was 0.96 and 0.97 for inter-reader reliability. Significant Spearman's correlations were found between segmented BML volume and WORMS BML scoring for tibial medial (0.75) and lateral (0.73) compartments, and for femoral medial (0.72) and lateral (0.88) compartments. Significant positive associations were found between weight-bearing pain and total femoral BML volume (p<0.003), but not total tibial BML (p<0.101)

Conclusion—We have documented a moderately strong correlation between a novel measurement method of femoral and tibial BML volume and semi-quantitative WORMS scores, providing evidence of criterion validity. The hypothesis that weight-bearing pain was associated with BML volume was confirmed for total femoral BML volume but not total tibial BML volume. The lack of association between tibial BML volume and pain requires further investigation.

Keywords

Osteoarthritis; Bone marrow lesions; Knee; Magnetic Resonance Imaging

Introduction

Bone marrow lesions (BML) are emerging as potentially important imaging biomarkers in knee osteoarthritis (OA), and have been associated with pain, cartilage thinning, meniscal pathology and bone attrition[1-4]. They are a potentially attractive proxy measure for documenting the natural history of symptomatic knee OA and for evaluating new therapies for prevention and treatment.

MRI offers detailed insight into OA pathology and is useful for evaluating BML[56, 7]. Clinically, BMLs are typically assessed with qualitative description, while in OA research the current standard is semi-quantitative scales, such as the Whole-Organ Magnetic Resonance Imaging Score (WORMS)[8] or the MRI Osteoarthritis Knee Score (MOAKS)[9]. Semi-quantitative methods use an ordinal score indicating the size of a BML relative to joint sub-regions. While this method has provided consistent scoring between studies, the cost is high due largely to reading times, radiology expertise required, and expert judgment required for each slice and sub-region. Further, semi-quantitative methods require a subjective decision as to size of BML and represent them with ordinal variables, losing some quantitative information.

An alternative is quantitative assessment. This approach may be more efficient, as the computer is programmed to automatically report the number of voxels included in the outlined object, and calculate the volume. It would also provide a quantitative outcome that may be more responsive to change in longitudinal studies. Further, once images have been screened by an MSK radiologist for alternative, potentially ominous diagnoses, a reader with less expertise but with knowledge of anatomy, pathology and radiological principles may be able to operate the software, after being trained by an experienced MSK radiologist. To date there have been a few reports of quantitative approaches to BML assessment and the methods described have been either time-consuming, involving either manual segmentation on each sagittal slice [10-12], have approximated BML volume using 2-dimensional measures of BML size or are 3-dimensional approximations of BML size [13-16]. Semi-automated methods that reliably and efficiently detect and measure BML volume would be an important advance. Such methods may be more responsive and because they are rapid, could be utilized in large OA trials and observational studies, such as the Osteoarthritis Initiative (OAI). The purpose of this paper was to validate a novel method of measuring OA-related BML volumes through comparison with the WORMS BML scores, and to demonstrate construct validity by confirming the new measurement method relates to weight-bearing pain, an association previously reported in the literature.

Materials and Methods

OAI study sample

We used the data from the OAI, a longitudinal multi-center cohort study of biomarkers and risk factors for the development and progression of knee OA. Of the 4,796 OAI participants,

1,380 represent a Progression subcohort containing 597 men and 793 women with frequent knee symptoms (defined as "pain, aching or stiffness in or around the knee on most days for at least one month" during the previous 12 months) and radiographic tibiofemoral knee OA (defined as definite tibiofemoral osteophytes: OARSI atlas grades 1-3[17], equivalent to K/L grade 2). General exclusion criteria (for all OAI participants) included rheumatoid or inflammatory arthritis, bilateral end-stage knee OA and MRI contraindications. A full description of study protocol, design, data overview and the datasets are available for public access at http://www.oai.ucsf.edu/. The study was HIPAA compliant and all subjects provided informed consent. The study protocol, amendments and informed consent documentation were reviewed and approved by the local institutional review boards.

OAI Central Image Assessments

The OAI has produced standardized measurements and interpretations of OAI radiographic and MRI images (central image assessments) from selected samples of participants in the Progression subcohort using documented methods and made them available online. Central image assessments were performed by qualified vendors selected by the OAI (http://www.oai.ucsf.edu/datarelease/ImageAssessments.asp#Datasets).

Study sample

For the current study, we used the central image assessment subsample of OAI participants whose MRI scans had been scored by an experienced MSK radiologist (AG) using WORMS (n=115). These subjects have been used in other studies and are described in other publications [18,19, 20]. In brief, people with baseline and 24-month follow-up MRIs and knee radiographs (OAI public datasets 0.2.1 and 1.2.1) who were at high risk of cartilage loss included those with baseline malalignment (> 2 degrees varus or valgus based on a hip-knee-ankle angle measured from full limb radiograph[21]), BML, definite osteophyte, and tibio-femoral joint space narrowing (JSN), as well as those with worsening JSN between the baseline and 12 month central reading[18, 20].

MRI and scoring methods

MR images were acquired at four OAI clinical centers using dedicated Siemens Trio 3T scanners (Trio, Siemens, Erlangen, Germany). Sagittal turbo spin echo fat-suppressed (TSE FS) ($0.357 \times 0.357 \times 3.0$ mm, TR 3200ms, TE 30ms) intermediate-weighted MRI were obtained. Details of the acquisition protocols and scoring systems have been previously published [22].

Semi-automated scoring of BMLs

An initial demarcation of relevant bone edges that did not require a skilled reader was performed by research assistants with no formal medical training and required approximately 5 minutes per knee (femur and tibia) and ranged from 1-8 minutes. A reader (CR) with knowledge of anatomy, pathology and radiological principles then used semi-automated software to segment the subchondral OA-related BMLs in the distal femur and proximal tibia. The reader undertook extensive training using an independent data set by two experienced radiologists (AG and CV). This included the differentiation of OA-related BML from similar signal due to partial volume averaging, insufficiency fracture, susceptibility artifact and other causes [23]. In addition, images in the study dataset where the reader was uncertain were reviewed with the radiologist.

The reader was blinded to subject ID and WORMS scoring. The software algorithm applied a local threshold based on the typical gray-level intensities in the surrounding bone of the femur (Figure 1a-c) and tibia and provided the reader with regions of hyperintense signal for

potential segmentation. Reader judgment was used to confirm, usually with 1 or 2 mouse clicks, the region of OA-related BML, defined as ill-delineated areas of increased signal intensity adjacent to subchondral bone (Figure 1c), and to reject irrelevant areas of increased signal.

Reliability was assessed on a random sample of 20 subjects. For intra-rater reliability readings were separated in time by 8 weeks to avoid recall bias. Inter-reader reliability was assessed on the same 20 cases by an experienced radiologist (CV).

Image Analysis

The primary outcome was total segmented volume of BMLs in the femoral and tibial medial and lateral compartments reported in cubic millimeters (mm^3), based on the number of hyperintense voxels selected. The division between medial and lateral compartments was identical to that used in WORMS (a line extending cranially from the lateral edge of femoral notch, with the trochlear groove defined as being on the medial side of the femur) [8]. Comparison of femoral and tibial BML volume as measured by the quantitative method was made with WORMS scoring of the same knee at the same time-point. Intra-rater and inter-rater reliability for WORMS scoring has been previously reported and is available on the OAI website (weighted kappa 0.66 and 0.94 respectively). In WORMS, the medial and lateral tibial and femoral compartments are divided into three sub-regions (anterior, central, posterior), and the tibia has an additional subspinous sub-region, comprised of the non-articulating portion of the tibial plateau beneath the tibial spines, and is viewed in the coronal plane. WORMS scoring then reports BML categorically for each sub-region, based on size of BML relative to the total sub-region (0=no BML, 1=<25% of the sub-region, 2=25-50% of the sub-region, 3=>50% of the sub-region) [8]. Since the quantitative scoring method reported values for both the entire medial compartment and entire lateral compartment (rather than sub-regions) of each bone, a summary measure of WORMS score was computed to give the following WORMS score for the hemi-compartment of each bone: WORMS $_{sum\ lateral}$ = lateral-anterior + lateral-central + lateral-posterior; WORMS $_{sum\ medial}$ = medial-anterior + medial-central + medial-posterior. (For the tibia, the subspinous region was included in the medial tibial compartment). Summation of WORMS scores was done to facilitate comparison of total BML in each hemi-compartment as measured by the semi-quantitative and quantitative methods. In order to limit the number of WORMS comparison categories and enhance clarity of presentation, the WORMS $_{summary}$ scores for each compartment were capped at three; i.e., if the sum was three (the highest possible WORMS score for a sub-region) or greater, the summary measure for the entire hemi-compartment was placed in the 3+ category. The 3+ category reflects relatively large and/or multiple BMLs in a femoral or tibial hemi-compartment; further categorization seemed unlikely to provide a significant gain in information.

In order to assess construct validity, we compared BML volumes with weight-bearing pain. Based on previous work [24-26], we used the WOMAC pain sub-scale and defined the primary outcome of knee pain dichotomously as moderate to severe pain (scores 2-4) on any of the 3 weight-bearing WOMAC pain questions (pain on walking, pain on climbing stairs, pain on standing), acquired at the same baseline OAI visit as the MRI scans. We also examined each specific weight-bearing item individually.

Statistical Analysis

Concurrent Validity—Descriptive data were summarized in tabular and graphical form to show quantitative BML volumes by each WORMS BML category in both medial and lateral compartments of the femur and tibia. Three analyses were carried out to examine concurrent validity of the quantitative measure against WORMS: 1) a non parametric (Spearman's)

correlation of WORMS categories and BML volume, 2) Kruskall-Wallis tests to compare median volumes across WORMS categories, and 3) Fisher's Exact test of the differences in proportions (estimated semi-automated BML = 0 vs. > 0) across the four categories of WORMS score. When the overall Krukall-Wallis test was significant, all pairwise comparisons were examined using the Wilcoxon rank sum test, with Bonferonni's correction to account for multiple comparisons.

Clinical Construct Validity—The associations between the weight-bearing pain sub-scale (as well as the individual weight-bearing pain items) and BML volumes (total femur and total tibia separately, and combined) were examined using the Wilcoxon rank sum test. For multivariable analysis, a square root transformation was used for each dependent variable and linear regression was used to examine the association between each dependent variable and pain, adjusted for age, sex, BMI, race (white vs non-white) and medial minimum joint space width.

Reliability—For both intra-reader and inter-reader reliability, the Shrout-Fleiss intraclass correlation coefficients (ICC) were calculated as measures of agreement. The ICC (2,1) correlation was calculated, used when all subjects are rated by the same raters who are assumed to be a random subset of all possible raters [27].

Results

Subject Characteristics

One hundred and fifteen subjects were included in the analysis. The sample was 84% white and 52% male. The distribution of baseline K/L grades was 34% subjects with grade 2, 55% with grade 3 and 7% with grade 4. Malalignment on long limb radiograph at the 12 month visit (>2 degrees of varus or valgus as measured by the HKA) was present in 78% of subjects. Table 1 shows subject characteristics.

Descriptive quantitative and semi quantitative BML data

Descriptive data for BML volumes in the femur and tibia by WORMS category are presented in Table 2 and 3 respectively. The software method required an average of 5 minutes per knee by the trained reader, not including the initial bone edge demarcation.

Concurrent Validity: Associations between quantitative method and WORMS scoring

Statistically significant positive associations (Figures 2 and 3, Tables 2 and 3) between the quantitative measure and WORMS summary scores were found for the medial and lateral compartments for both femur and tibia. Kruskal-Wallis tests of the differences in medians across the four categories of WORMS score was significant for the medial and lateral compartments of both bones (p<0.001 femur, p<0.001 tibia). All pairwise comparisons were examined using the Wilcoxon rank sum test and were statistically significant (using Bonferonni's correction for multiple comparisons, p<0.008), with the exception of medial femur WORMS 2 vs. 3+, lateral tibia WORMS 2 vs. 3+, and medial tibia WORMS 1 vs. 2. Fisher's Exact test of the differences in proportions (any BML vs. no BML) across the four categories of WORMS score was significant for the medial and lateral compartments of both femur and tibia. Spearman's correlation between BML volume and WORMS score by femoral lateral compartment was 0.88, femoral medial compartment 0.72, tibial lateral 0.73 and tibial medial 0.75 (all p<0.0001, 95% CIs shown in tables).

Clinical Construct Validity: Associations between quantitative method and WOMAC pain scores

Subjects with weight-bearing knee pain had greater median BML volume than those without weight-bearing knee pain. Results of the association between BML volume and weight-bearing pain sub-scale, adjusted for age, sex, BMI, race, and medial minimum joint space width, are presented in Table 4 (crude associations were very similar; statistical significance of each association did not change). In the examination of each specific weight-bearing item, we found an association with greater BML volume and pain during stair-climbing for both the femur (median BML 0.00 mm^3 no pain vs. 212.6 mm^3 pain, p=0.02) and tibia (141.1 mm^3 no pain vs. 380.4 mm^3 pain, p=0.03), while neither femoral nor tibial BML volume was associated with pain during walking or standing (data not shown).

Reliability

The ICC for intra-reader reliability was 0.96 (95% CI, 0.90-0.98) and 0.97 (95% CI, 0.93-0.99) for inter-reader reliability, as assessed in the random sample of 20 subjects.

Discussion

This study provides evidence for the validity and reliability of a novel semi-automated method to measure the volume of BMLs in knee OA patients. Evidence of validity was demonstrated through general agreement between BML volume and WORMS BML categories - a current semi-quantitative gold standard in OA imaging research, and by an association between femoral BML volume and weight-bearing pain. Interestingly, no association between tibial BML volume and the weight-bearing pain sub-scale was demonstrated.

While there was broad agreement between BML volume and WORMS BML categories, part of the unexplained variation may be due to inherent methodological differences. Both may provide useful but somewhat different information, similar to the radiographic measures of joint space width and K/L grade. While WORMS provides an ordinal score that offers researchers an immediate idea of size of BML that may be a gauge for severity of disease, quantitative measures of BML provide a continuous score that is not immediately interpretable by researchers. However the continuous score has the potential to be a more sensitive, responsive instrument for monitoring longitudinal change. On the other hand, it may pick up changes that are not clinically meaningful. Assessing the responsiveness to clinically meaningful change is an important next research goal.

The new method is also efficient – in this sample of subjects with established radiographic knee OA and a high prevalence of BML, the average reading time was 5 minutes per knee. The measurements were performed by one trained reader (CR) with anatomical knowledge and training from expert radiologists. One of the significant challenges of previous methods to measure BML volumetrically has been the time required to segment each slice manually[10-12]. The introduction of a semi-automated rapid, valid and reliable quantitative measurement for BML has implications for study costs and power.

Most volumetrically measured BMLs aligned with WORMS ordinal categories, although there was some overlap in BML volumes across WORMS categories (Figures 2 and 3). There are several probable reasons for some differences between the two systems. WORMS evaluates BML size relative to the overall region (e.g., > 50% of subregion) while the BML volume is an absolute value. The new measure is quantitative and reported as a continuous variable (mm^3) while WORMS is ordinal. Since there is some subjectivity in assigning BMLs to WORMS categories, especially at the boundaries of categories, some overlap in volume measurements between adjacent categories of WORMS is not surprising. There is

also some subjectivity and reader judgment in both methods. In both cases, the reader must be able to distinguish subchondral OA-related BML from other types of BML (e.g., traumatic) and other causes of similar signal alterations. Disagreement for the most part was centered around small ill-defined BML's that were not scored by the volumetric method, but that WORMS central reading had scored as BML. This was more evident in the tibia and may have been due to methodological differences between the instruments. WORMS uses a coronal image to better visualize the subspinous region, the quantitative method did not. BMLs in this region are challenging to segment as some are related to cruciate ligament pathology, some extend into the adjacent medial compartment and associated with cartilage loss and some are related to insertional cysts [28]. Some of the discordance may have been also been due to reader disagreement or, small ill-defined BMLs may not have been detected by the thresholding algorithm. Adjustments to the thresholding algorithm might have improved performance. Further work is being undertaken to examine modifications to the method in evaluation of this region. Achieving higher sensitivity for small BMLs (without sacrificing specificity) is a key goal of ongoing research. Nevertheless, despite the different methodologies, there was general agreement between the two instruments. It is also important to note that the distinction of a BML due to OA and one resulting from another process, such as trauma, is critical to the care of individual patients, but may be less critical in research studies that aim to identify and/or validate relationships across large samples.

Past studies have provided a moderate level of evidence for the relationship between BML and pain [29], with most [1, 24, 30-33] but not all [34-36] documenting an association. One recent study showed a relationship with pain on stair climbing but not on total WOMAC pain or WOMAC walking [26]. Most studies have used a binary pain outcome (yes/no) [31, 33, 37], a visual analogue scale [32], or the WOMAC pain scale [1, 35, 36]. Recently two studies have used the WOMAC weight-bearing sub-scale [24], or individual weight-bearing WOMAC items [26].

Our data for the weight-bearing pain sub-scale revealed a relationship with femoral but not tibial BML volume. The lack of association between tibial BML volume and pain has not been previously reported and requires further investigation. This is the first study to use a quantitative measure of BML and correlate it with pain, and we did not examine a threshold size or location, which may have attenuated the relationship in the tibia. Other studies have used a summed BML score (ordinal scores summed across whole knee) [1, 26] or maximal score (from region with highest BML score)[24], any or large BML[33], or compartment-specific BML[31, 32]. Of interest, in examining our question-specific data, pain on climbing stairs was significantly related to both femoral and tibial BML volume, while there was no relationship with BML in either bone for standing pain or walking pain, consistent with the recent study by Ip et al (using a 4-point ordinal score for BML at 6 sub-regions) [26]. It is biologically plausible that higher mechanical loads may play a role in the relationship between BML and pain and requires further study.

Strengths of this study included the blinded readings, the standardized methods used for data acquisition and the high intra and inter-reader reliability. We measured BML volume in a sample with established knee OA in which over half the sample had BML representing a wide variety in size and shape. The semi-automated methodology was able to identify and segment these BML in most cases, and minimal manual correction was required. Even given the high prevalence of BML, reading time remained relatively quick. In studies of incidence or in earlier stages of OA, reading times will likely be quicker given the lower prevalence of BML.

This study had several limitations. The images were acquired on MRI scanners using a 3T TSE FS IM sequence which provided excellent visualization of BML. We did not examine other MRI sequences, where performance may vary. We also did not examine the patellar

side of the patella-femoral joint – a common site for BML in OA. We did not compare the new volumetric measurements to other semi-quantitative methods such as the Boston Leeds Osteoarthritis Knee Score (BLOKS) [38] or MOAKS, since only WORMS scores were available from OAI central imaging assessments. WORMS and BLOKS have been shown to be relatively comparable, having high reliability[20, 38, 39] for most BML scores. In the cross-sectional comparisons for BML, the two methods gave similar results for prevalence and severity of BMLs, while in longitudinal comparison BML scoring in WORMS was preferable in that it better predicted later cartilage loss, was easier to score and did not include potentially extraneous measures [39]. Another potential limitation is that the quantitative method as presented scores the entire medial and lateral hemi-compartments, while semi-quantitative methods such as WORMS report on sub-regions. Sub-regional assessment allows for measurement of clinically important areas.

However, work is underway to have the computer automatically divide the knee into clinically important sub-regions that would automatically produce quantitative measures for both sub-regional and total compartment on request. Again, there are no perceived barriers to its implementation. Well-defined cysts were excluded from the analysis, though cysts that were part of an ill-defined BML were not. While it was possible for the reader to exclude these cystic regions, a methodological decision, based partly on efficiency and partly on biologic rationale was made to include them - this likely resulted in underestimation of the correlation with WORMS BML scores which excludes all cysts from BML scores. Lastly, the quantitative method did not provide information on the number of lesions per knee, though this could readily be included in the future.

With regard to the relationship with pain, we are reporting on the validation of a new methodology and caution should be applied in drawing any definitive conclusions about potentially clinically relevant associations. In this case, we can't be sure that using total BML volume (rather than sub-regional), not defining a threshold size of BML, or some inherent weakness of the method has not obscured the relationship with pain.

Several other studies report on the quantitative measurement of BML using manual segmentation. Schmid et al [12] used a method that involved manual segmentation of the BML on each slice and subsequent multiplication of the segmented areas in each slice by the slice thickness – a method that is time consuming. In an attempt to minimize the observer-dependent nature of the BML volume quantification process, Mayerhoefer[11] reported on a computer-assisted model to automatically calculate bone marrow volume at the knee, which also used a thresholding technique. Times were not reported in this article, though it has been reported by others as time-consuming [10]. Roemer et al recently reported on volumetric measurement of BML using a modified version of the method originally suggested by Schmid et al [10, 12]. BML volume was calculated from the segmented images by means of manual segmentation of the lesion area on each sagittal slice and by multiplication of the section thickness plus the inter-slice gap. In the modified method, it was not necessary to manually multiply the area segmented on each slice by the slice thickness as this was an automatic function integrated in the segmentation software, however reading times were not reported. The paper compared manual volumetric measures of BML (as well as WORMS), and reported that volumetric (and WORMS scores) of subchondral BMLs in OA can be performed reliably on proton density-weighted fat suppressed and T1-weighted fat-suppressed contrast-enhanced sequences.

Further research with this new method is underway. This includes longitudinal assessment of change in BML size (responsiveness), and the relationship between BML volumes (and changes in volume) to clinically important endpoints. Work is also continuing on packaging semi-automated quantitative measures of BML volume with similar semi-automated

quantitative measures of cartilage[40], osteophyte and meniscus into a rapid, quantitative measure for whole organ MRI assessment of the knee.

In conclusion, we have provided evidence for the validity of an efficient, quantitative method of BML volume in knee OA subjects using MRI. This technique could provide a rapid, valid quantitative measure of BML, making it feasible to assess a large number of knees in a short period of time.

Acknowledgments

Appendix 1

Causes of subchondral signal alterations of the knee joint, not due to osteoarthritis
(a) Subchondral Insufficiency Fracture - Sagittal IW fat-suppressed MRI shows a subtle subchondral irregular hypointense line (arrowheads) of the medial femoral condyle which represents subchondral insufficiency fracture surrounded by an extensive hyperintensity bone marrow edema (arrows). Also there is a large heterogeneous hyperintensity of the medial tibial plateau (thin arrows) typical for osteoporosis in this 59-year-old woman. Note there is a small joint effusion.
(b) Traumatic Bone Marrow Secondary to Transient Lateral Subluxation of the Patella - Sagittal IW fat-suppressed MRI shows moderate hyperintensity of the lateral femoral condyle (arrows) distant from the subchondal bone and typical for traumatic bone marrow secondary to transient lateral subluxation of the patella. Note there is also a grade 1 medial collateral ligament sprain

References

1. Zhang Y, Nevitt M, Niu J, Lewis C, Torner J, Guermazi A, et al. Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. Arthritis Rheum. 2011; 63:691–699. [PubMed: 21360498]

2. Roemer FW, Neogi T, Nevitt MC, Felson DT, Zhu Y, Zhang Y, et al. Subchondral bone marrow lesions are highly associated with, and predict subchondral bone attrition longitudinally: the MOST study. Osteoarthritis Cartilage. 2010; 18:47–53. [PubMed: 19769930]

3. Dore D, Martens A, Quinn S, Ding C, Winzenberg T, Zhai G, et al. Bone marrow lesions predict site-specific cartilage defect development and volume loss: a prospective study in older adults. Arthritis Res Ther. 2010; 12:R222. [PubMed: 21190554]

4. Englund M, Guermazi A, Roemer FW, Yang M, Zhang Y, Nevitt MC, et al. Meniscal pathology on MRI increases the risk for both incident and enlarging subchondral bone marrow lesions of the knee: the MOST Study. Ann Rheum Dis. 2010; 69:1796–1802. [PubMed: 20421344]

5. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Li L, et al. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. Osteoarthritis Cartilage. 2011; 19:557–588. [PubMed: 21396463]

6. Roemer FW, Hunter DJ, Guermazi A. MRI-based semiquantitative assessment of subchondral bone marrow lesions in osteoarthritis research. Osteoarthritis Cartilage. 2009; 17:414–415. author reply 416-417. [PubMed: 18948039]

7. Roemer FW, Eckstein F, Guermazi A. Magnetic resonance imaging-based semiquantitative and quantitative assessment in osteoarthritis. Rheum Dis Clin North Am. 2009; 35:521–555. [PubMed: 19931802]

8. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage. 2004; 12:177–190. [PubMed: 14972335]

9. 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. [PubMed: 21645627]

10. Roemer FW, Khrad H, Hayashi D, Jara H, Ozonoff A, Fotinos-Hoyer AK, et al. Volumetric and semiquantitative assessment of MRI-detected subchondral bone marrow lesions in knee osteoarthritis: a comparison of contrast-enhanced and non-enhanced imaging. Osteoarthritis Cartilage. 2010; 18:1062–1066. [PubMed: 20472082]

11. Mayerhoefer ME, Breitenseher MJ, Kramer J, Aigner N, Norden C, Hofmann S. STIR vs. T1-weighted fat-suppressed gadolinium-enhanced MRI of bone marrow edema of the knee: computer-assisted quantitative comparison and influence of injected contrast media volume and acquisition parameters. J Magn Reson Imaging. 2005; 22:788–793. [PubMed: 16270290]

12. Schmid MR, Hodler J, Vienne P, Binkert CA, Zanetti M. Bone marrow abnormalities of foot and ankle: STIR versus T1-weighted contrast-enhanced fat-suppressed spin-echo MR imaging. Radiology. 2002; 224:463–469. [PubMed: 12147843]

13. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Abram F, Choquette D, Haraoui B, 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; 67:683–688. [PubMed: 17728333]

14. Driban JB, Lo GH, Lee JY, Ward RJ, Miller E, Pang J, 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. [PubMed: 21961433]

15. Bining HJ, Santos R, Andrews G, Forster BB. Can T2 relaxation values and color maps be used to detect chondral damage utilizing subchondral bone marrow edema as a marker? Skeletal Radiol. 2009; 38:459–465. [PubMed: 19089419]

16. Roemer FW, Bohndorf K. Long-term osseous sequelae after acute trauma of the knee joint evaluated by MRI. Skeletal Radiol. 2002; 31:615–623. [PubMed: 12395272]

17. Altman RD, Hochberg M, Murphy WA Jr, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage. 1995; 3(A):3–70. [PubMed: 8581752]

18. Lo GH, Hunter DJ, Zhang Y, McLennan CE, Lavalley MP, Kiel DP, et al. Bone marrow lesions in the knee are associated with increased local bone density. Arthritis Rheum. 2005; 52:2814–2821. [PubMed: 16145676]

19. OAI Central Semi-Quantitative Reading of Knee MRIs for OA-Released Structural Change (Boston Imaging Core Lab). http://oai.epi-ucsf.org/datarelease/ImageAssessments.asp

20. 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; 18:1393–1401. [PubMed: 20816979]

21. Felson DT, Cooke TD, Niu J, Goggins J, Choi J, Yu J, et al. Can anatomic alignment measured from a knee radiograph substitute for mechanical alignment from full limb films? Osteoarthritis Cartilage. 2009

22. 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; 16:1433–1441. [PubMed: 18786841]

23. Roemer FW, Frobell R, Hunter DJ, Crema MD, Fischer W, Bohndorf K, et al. MRI-detected subchondral bone marrow signal alterations of the knee joint: terminology, imaging appearance, relevance and radiological differential diagnosis. Osteoarthritis Cartilage. 2009; 17:1115–1131. [PubMed: 19358902]

24. Lo GH, McAlindon TE, Niu J, Zhang Y, Beals C, Dabrowski C, 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

25. Stratford PW, Kennedy DM, Woodhouse LJ, Spadoni GF. Measurement properties of the WOMAC LK 3.1 pain scale. Osteoarthritis Cartilage. 2007; 15:266–272. [PubMed: 17046290]

26. Ip S, Sayre EC, Guermazi A, Nicolaou S, Wong H, Thorne A, et al. Frequency of bone marrow lesions and association with pain severity: results from a population-based symptomatic knee cohort. J Rheumatol. 2011; 38:1079–1085. [PubMed: 21362771]

27. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428. [PubMed: 18839484]

28. Hernandez-Molina G, Guermazi A, Niu J, Gale D, Goggins J, Amin S, et al. Central bone marrow lesions in symptomatic knee osteoarthritis and their relationship to anterior cruciate ligament tears and cartilage loss. Arthritis Rheum. 2008; 58:130–136. [PubMed: 18163483]

29. Yusuf E, Kortekaas MC, Watt I, Huizinga TW, Kloppenburg M. Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. Ann Rheum Dis. 2011; 70:60–67. [PubMed: 20829200]

30. Torres L, Dunlop DD, Peterfy C, Guermazi A, Prasad P, Hayes KW, et al. The relationship between specific tissue lesions and pain severity in persons with knee osteoarthritis. Osteoarthritis Cartilage. 2006; 14:1033–1040. [PubMed: 16713310]

31. Hayes CW, Jamadar DA, Welch GW, Jannausch ML, Lachance LL, Capul DC, et al. Osteoarthritis of the knee: comparison of MR imaging findings with radiographic severity measurements and pain in middle-aged women. Radiology. 2005; 237:998–1007. [PubMed: 16251398]

32. Sowers MF, Hayes C, Jamadar D, Capul D, Lachance L, Jannausch M, et al. Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis. Osteoarthritis Cartilage. 2003; 11:387–393. [PubMed: 12801478]

33. Felson DT, Chaisson CE, Hill CL, Totterman SM, Gale ME, Skinner KM, et al. The association of bone marrow lesions with pain in knee osteoarthritis. Ann Intern Med. 2001; 134:541–549. [PubMed: 11281736]

34. Kornaat PR, Bloem JL, Ceulemans RY, Riyazi N, Rosendaal FR, Nelissen RG, et al. Osteoarthritis of the knee: association between clinical features and MR imaging findings. Radiology. 2006; 239:811–817. [PubMed: 16714463]

35. Phan CM, Link TM, Blumenkrantz G, Dunn TC, Ries MD, Steinbach LS, et al. MR imaging findings in the follow-up of patients with different stages of knee osteoarthritis and the correlation with clinical symptoms. Eur Radiol. 2006; 16:608–618. [PubMed: 16222533]

36. Link TM, Steinbach LS, Ghosh S, Ries M, Lu Y, Lane N, et al. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. Radiology. 2003; 226:373–381. [PubMed: 12563128]

37. Kornaat PR, Kloppenburg M, Sharma R, Botha-Scheepers SA, Le Graverand MP, Coene LN, et al. Bone marrow edema-like lesions change in volume in the majority of patients with osteoarthritis; associations with clinical features. Eur Radiol. 2007; 17:3073–3078. [PubMed: 17823802]

38. 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; 67:206–211. [PubMed: 17472995]

39. Felson DT, Lynch J, Guermazi A, Roemer FW, Niu J, McAlindon T, 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; 18:1402–1407. [PubMed: 20851202]

40. Iranpour-Boroujeni T, Watanabe A, Bashtar R, Yoshioka H, Duryea J. Quantification of cartilage loss in local regions of knee joints using semi-automated segmentation software: analysis of longitudinal data from the Osteoarthritis Initiative (OAI). Osteoarthritis Cartilage. 2010

(a)

(b)

(c)
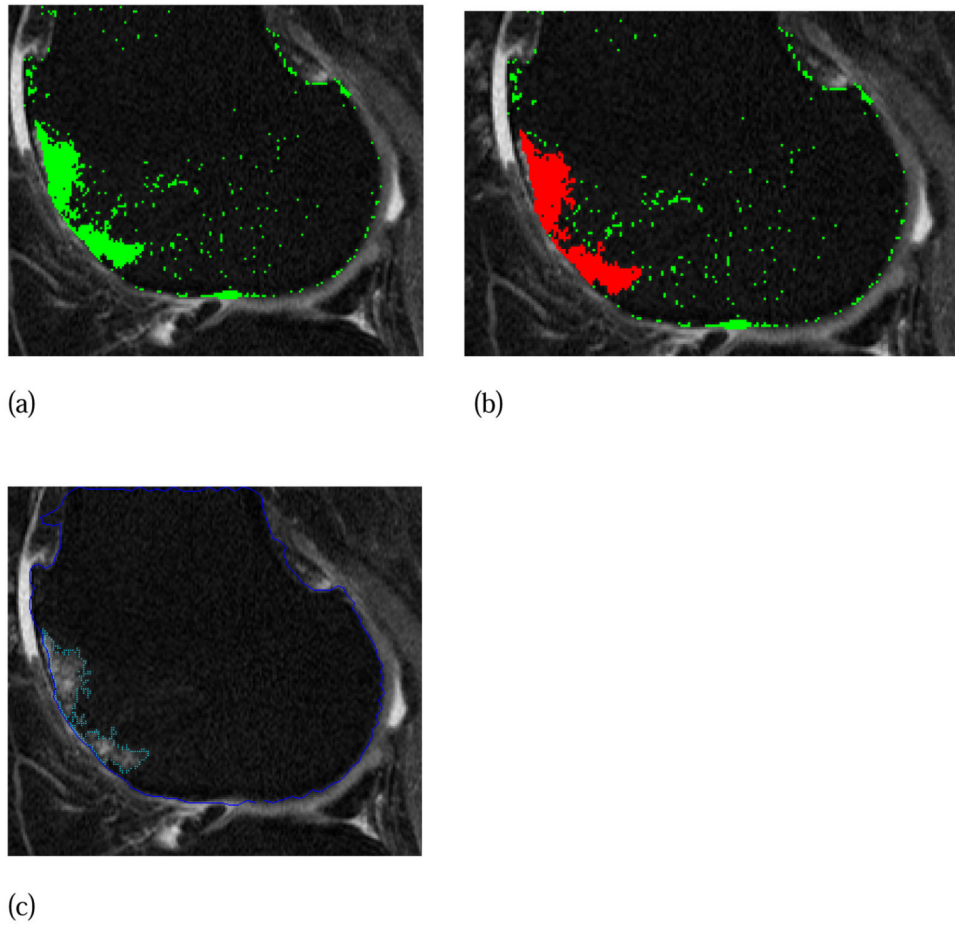
Figure 1. BML semi-automated segmentation. (a) software presents hyperintense areas in green based on gray-scale thresholding, (b) reader selects areas of subchondral OA related BML-green changes to red (2 mouse clicks in this case), (c) appearance of final segmented BML
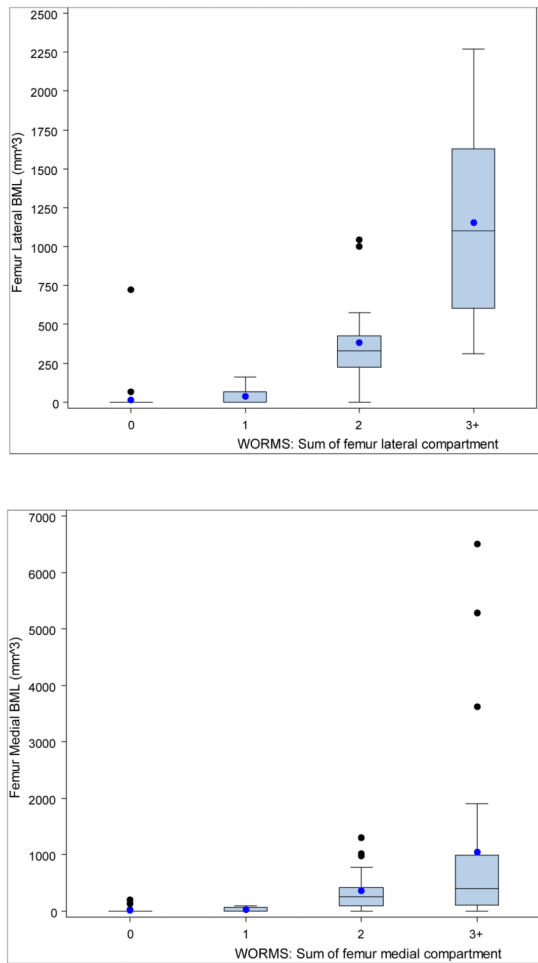
Figure 2. Box and whisker plots showing mean, median, lower and upper quartiles, and outliers of BML volume from semi-automated quantitative assessment, by WORMS scores, for femoral (a) lateral compartment, and (b) medial compartment
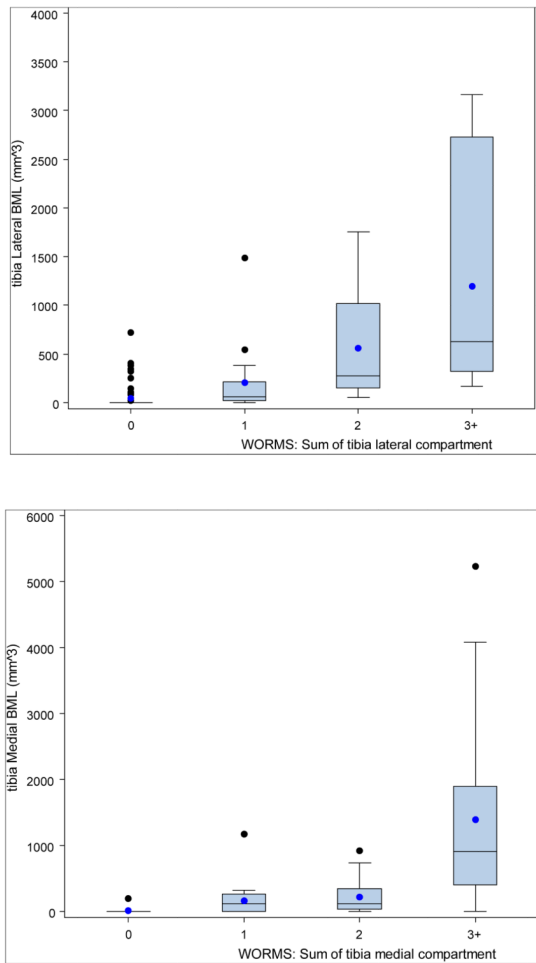
Figure 3. Box and whisker plots showing mean, median, lower and upper quartiles, and outliers of BML volume from semi-automated quantitative assessment, by WORMS scores, for tibial (a) lateral compartment, and (b) medial compartment

Table 1

Subject Characteristics

	N (%)
Sex	
Male	60 (52%)
Female	55 (48%)
Age Group	
45-49	16 (14%)
50-59	31 (27%)
60-69	38 (33%)
70-79	30 (26%)
Race	
White	97 (84%)
Non-White	18 (16%)
Kellgren and Lawrence grade	
0	2 (2%)
1	3 (3%)
2	39 (34%)
3	63 (55%)
4	8 (7%)
Alignment	
Neither	25 (22%)
Varus (>2 degrees)	27 (24%)
Valgus (> 2 degrees)	62 (54%)

Table 2

Femur BML Volume vs WORMS scores

(a) Lateral compartment

	0	1	2	3+	p-value
N	67	14	17	16	
Mean	11.8	37.7	381.1	1153.7	
Std Dev	88.7	55.2	285.0	615.0	
Minimum	0.00	0.00	0.00	312.00	
Median	0.0	0.0	331.9	1100.0	<0.001[1]
Maximum	723.8	160.2	1043.4	2268.8	
n(%) = 0	65 (97%)	8 (57%)	1 (6%)	0 (0%)	
n(%) >0	2 (3%)	6 (43%)	16 (94%)	16 (100%)	<0.001[2]

[1]Kruskill-Wallis. All pairwise comparisons significant.

[2]Fisher's Exact test

	Spearman Correlation		
	Coefficient	**95% CI**	**p-value**
	0.88	0.83, 0.91	<.0001

(b) Medial compartment

	0	1	2	3+	p-value
N	45	25	21	24	
Mean	8.0	26.1	355.2	1039.8	
Std Dev	35.5	35.1	362.1	1709.0	
Minimum	0.00	0.00	0.00	0.00	
Median	0.00	0.00	248.14	405.86	<0.001[1]
Maximum	199.9	91.4	1299.2	6502.5	
n(%) = 0	42 (93%)	15 (60%)	3 (14%)	4 (17%)	
n(%) >0	3 (7%)	10 (40%)	18 (86%)	20 (83%)	<0.001[2]

[1]Kruskill-Wallis. All pairwise comparisons significant.

[2]Fisher's Exact test

(a) Lateral compartment				
0	**1**	**2**	**3+**	**p-value**
	Spearman Correlation			
	Coefficient	**95% CI**	**p-value**	
	0.72	0.62, 0.80	<.0001	

Table 3

Tibia BML Volume vs WORMS scores

(a) Lateral compartment

	0	1	2	3+	p-value
N	80	17	11	7	
Mean	41.7	206.4	554.8	1193.3	
Std Dev	122.1	364.1	555.8	1224.5	
Minimum	0.00	0.00	55.82	167.09	
Median	0.00	56.6	277.5	624.0	<0.001[1]
Maximum	722.2	1485.0	1754.9	3163.2	
n(%) = 0	67 (84%)	3 (18%)	0 (0%)	0 (0%)	
n(%) >0	13 (16%)	14 (82%)	11 (100%)	7 (100%)	<0.001[2]

[1]Kruskill-Wallis. All pairwise comparisons significant.

[2]Fisher's Exact test

Spearman Correlation

Coefficient	95% CI	p-value
0.73	0.63, 0.80	<.0001

(b) Medial compartment

	0	1	2	3+	p-value
N	21	28	22	43	
Mean	9.6	157.6	217.9	1394.5	
Std Dev	41.5	231.3	262.0	1393.5	
Minimum	0.00	0.00	0.00	0.00	
Median	0.00	110.3	113.7	904.6	<0.001[1]
Maximum	190.4	1168.8	916.4	5226.6	
n(%) = 0	19 (90%)	10 (36%)	5 (23%)	1 (2%)	
n(%) >0	2 (10%)	18 (64%)	17 (77%)	42 (98%)	<0.001[2]

[1]Kruskill-Wallis. All pairwise comparisons significant.

[2]Fisher's Exact test

(a) Lateral compartment				
0	**1**	**2**	**3+**	**p-value**
	Spearman Correlation			
	Coefficient	**95% CI**	**p-value**	
	0.75	0.65, 0.82	<.0001	

Table 4

Associations of weight-bearing pain with BML volumes and WORMS scores

BML Measure		No weight-bearing pain	Weight-bearing pain	p-value[1]
Total BML (mm^3)	n mean (sd) median 25th, 75th percentile	54 1004.1 (1708.8) 416.6 55.1, 1325.2	61 1576.1 (2030.3) 864.9 377.8, 1649.1	0.007
Total Femur BML (mm^3)	n mean (sd) median 25th, 75th percentile	54 297.6 (521.9) 0.0 0.0, 443.1	61 714.9 (1210.6) 308.2 0.0, 772.0	0.003
Total Tibia BML (mm^3)	n mean (sd) median 25th, 75th percentile	54 706.4 (1370.3) 270.1 0.0, 685.9	61 861.2 (1211.1) 367.4 115.1, 1056.8	0.101

[1] adjusted for age, sex, BMI, race, and medial minimum joint space width