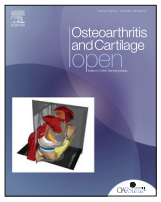
Osteoarthritis and Cartilage Open 6 (2024) 100465

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131



Approaches to optimize analyses of multidimensional ordinal MRI data in osteoarthritis research: A perspective

Jamie E. Collins [a,*], Frank W. Roemer [b,c], Ali Guermazi [b,d]

[a] Orthopaedics and Arthritis Center of Outcomes Research, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, BTM Suite 5016, Boston, MA, 02115, USA
[b] Quantitative Imaging Center, Department of Radiology, Boston University School of Medicine, 820 Harrison Avenue, FGH Building, 4th Floor, Boston, MA, 02118, USA
[c] Department of Radiology, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Universitätsklinikum Erlangen, Maximiliansplatz 3, 91054 Erlangen, Germany
[d] Department of Radiology, VA Boston Healthcare System, 1400 VFW Parkway, Suite 1B105, West Roxbury, MA, 02132, USA

ARTICLE INFO

Handling Editor: Professor H Madry

Keywords:
MRI
Clinical trials
Outcomes research

ABSTRACT

Objective: Knee osteoarthritis (OA) is a disease of the whole joint involving multiple tissue types. MRI-based semi-quantitative (SQ) scoring of knee OA is a method to perform multi-tissue joint assessment and has been shown to be a valid and reliable way to measure structural multi-tissue involvement and progression of the disease. While recent work has described how SQ scoring may be used for clinical trial enrichment and disease phenotyping in OA, less guidance is available for how these parameters may be used to assess study outcomes.
Design: Here we present recommendations for summarizing disease progression within specific tissue types. We illustrate how various methods may be used to quantify longitudinal change using SQ scoring and review examples from the literature.
Results: Approaches to quantify longitudinal change across subregions include the count of number of subregions, delta-subregion, delta-sum, and maximum grade changes. Careful attention should be paid to features that may fluctuate, such as bone marrow lesions, or with certain interventions, for example pharmacologic interventions with anticipated cartilage anabolic effects. The statistical approach must align with the nature of the outcome.
Conclusions: SQ scoring presents a way to understand disease progression across the whole joint. As OA is increasingly recognized as a heterogeneous disease with different phenotypes a better understanding of longitudinal progression across tissue types may present an opportunity to match study outcome to patient phenotype or to treatment mechanism of action.

1. Introduction

Osteoarthritis (OA) affects 300 million adults worldwide; with an aging population and the growing obesity epidemic, the prevalence is increasing [1,2]. OA was recently ranked as the 11th highest contributor to global disability of 291 conditions studied [3]. While traditionally considered a disease of aging, recent data suggest that over half of knee OA patients in the US are under the age of 65 [2]. Despite the clinical and economic burden of OA, no disease-modifying OA drugs (DMOADs) are currently available and current treatments are only modestly efficacious for symptoms and usually with side effects, serious in some cases [4,5].

One of the reasons for the failure of DMOAD trials is the primary use of radiography to determine study eligibility and to measure structural endpoints [6,7]. Many exclusionary conditions are X-ray-occult, tissue damage varies widely in supposedly eligible joints defined as Kellgren-Lawrence grades 2 and 3, and joint space loss as an outcome measure is a surrogate of damage in multiple tissues including cartilage and meniscus [8,9]. Moreover, posteroanterior radiography does not show synovitis, joint effusion and bone marrow lesions, features that are associated with pain and structural progression. Longitudinal reproducibility of X-ray positioning is challenging and often problematic [10]. Knee OA is a disease involving all tissues of the affected joint, including cartilage, bone, ligaments, menisci, muscles and synovium. The progression of OA is a complex process involving inflammatory, mechanical, genetic and metabolic factors and prediction of the disease course is challenging [11]. MRI-based semi-quantitative (SQ) scoring of knee OA is a method to perform multi-tissue joint assessment and has been shown to be a valid and reliable way to measure structural multi-tissue

* Corresponding author. Orthopedics and Arthritis Center for Outcomes Research, Brigham and Women's Hospital, 75 Francis Street, BTM 5-016, Boston, MA, 02115, USA.
E-mail address: Jcollins13@bwh.harvard.edu (J.E. Collins).

https://doi.org/10.1016/j.ocarto.2024.100465
Received 17 January 2024; Accepted 22 March 2024

involvement and progression of knee osteoarthritis [12,13]. Both cross-sectional presence and changes in features assessed by SQ imaging have been shown to be associated with subsequent progression [14–16].

The two most widely used SQ scoring systems for knee OA, MOAKS (MRI Osteoarthritis Knee Score) and WORMS (Whole-Organ Magnetic Resonance Imaging Score), rely on ordinal ratings of knee features by expert readers [14,17]. While guidance is available to describe how SQ scoring may be applied and can be used for clinical trial enrichment, less information is available on how these parameters should be used to assess outcomes [12,13,18–21].

Definitions of longitudinal change using MOAKS sores have been suggested previously, but suggestions for analytic strategies how to best leverage these rich multi-dimensional datasets are largely missing [21]. Without a robust analytic approach to best synthesize these ratings we may be under-utilizing the richness of the SQ MRI-based data. Here we describe how MOAKS can be used to quantify longitudinal change in knee OA and highlight its potential as an outcome measure in addition to other commonly used metrics such as quantitative approaches based on image segmentation.

2. Scoring methodology

The MOAKS system describes key pathoanatomic features of the joint, including cartilage damage (both in surface area extent and in full-thickness loss), meniscus damage, osteophytes, bone marrow lesions (BMLs), Hoffa-synovitis, effusion-synovitis and others [14]. The knee joint is divided into subregions (e.g., 14 for cartilage and bone marrow lesions) and locations (e.g., 10 locations for osteophyte assessment), and each is scored for a given feature on an ordinal scale ranging from 0 to 3. The MOAKS system and its application has been described previously and a brief overview is provided in the appendix [12,14].

2.1. Quantifying longitudinal change

Different approaches have been suggested to quantify longitudinal change in cartilage, osteophytes, and BMLs:

Count of number of subregions: Worsening across subregions is quantified by the number of subregions with a worse (higher) score at follow-up as compared to baseline, by the change in the number of subregions affected (score = 0 at baseline and >0 at follow-up), and by the number of subregions with true worsening of damage (i.e., score of >0 at baseline that shows worsening at follow-up). Improvement across subregions is quantified as the number of subregions with improvement from baseline to follow-up (i.e., lower score at follow-up as compared to baseline).

Delta-subregion (Delta-SR): building on the approach of change in number of subregions affected, the delta-SR approach considers worsening and improvement simultaneously. This is particularly relevant for bone marrow lesions that may show fluctuation in both directions over time in multiple subregions (Fig. 1). Both the number of subregions with worsening and the number of subregions with improvement are calculated, and the delta-SR change is calculated as the number of subregions with worsening minus the number of subregions with improvement.

Delta-sum: Add the absolute score across all subregions separately for each timepoint. Longitudinal change (delta-sum) can be quantified by taking the difference between follow-up and baseline. This approach has the drawback that the same delta-sum score may reflect minor changes in multiple subregions or large change in only few subregions.

Maximum grade: Calculate the change in each subregion. Longitudinal change is quantified as the maximum change across all subregions.

2.2. Additional considerations

- *Within-grade changes:* within-grade changes are changes that do not fulfill the definition of a full-grade change but do represent definite SQ visual change (e.g., within a subregion, the extent of cartilage area affected increasing from 25% to 60% would not meet the definition of full grade change; both timepoints would be coded as a grade of 2: 10–75% of subregion affected. Such a change would meet the definition of within-grade change). These changes have been shown to be clinically meaningful, and including such changes in SQ assessment of longitudinal change increases sensitivity to change [22]. Validity has been confirmed for longitudinal cartilage assessment showing strong associations with quantitative cartilage loss based on 3D segmentation approaches [23]. An example of calculating change metrics with and without including within-grade changes is provided in Table 1. An illustrative example of within-grade BML change is shown in Fig. 2.
- *Compartment specific vs. whole knee:* Cartilage morphology, BMLs, and osteophytes may be summarized over the whole joint, or separately for the medial femorotibial joint (MFTJ), lateral femorotibial joint (LFTJ), and patella-femoral joint (PFJ) [24]. For example, by counting the number of subregions with BML for the whole knee (range: 0 to 14) or in each compartment separately (range: 0 to 5 in the LFTJ and MFTJ, 0 to 4 in the PFJ) or by calculating the delta-sum across all SR in the whole knee or in each compartment separately, as shown in Roemer et al. [25]. Analyzing the data separately by compartment or for the whole joint may have implications for the statistical approach (see statistical considerations below).
- *Additional consideration for features that may fluctuate:* Some joint features, including BMLs, Hoffa-synovitis, and effusion-synovitis, may show both worsening and improvement. For example, a participant

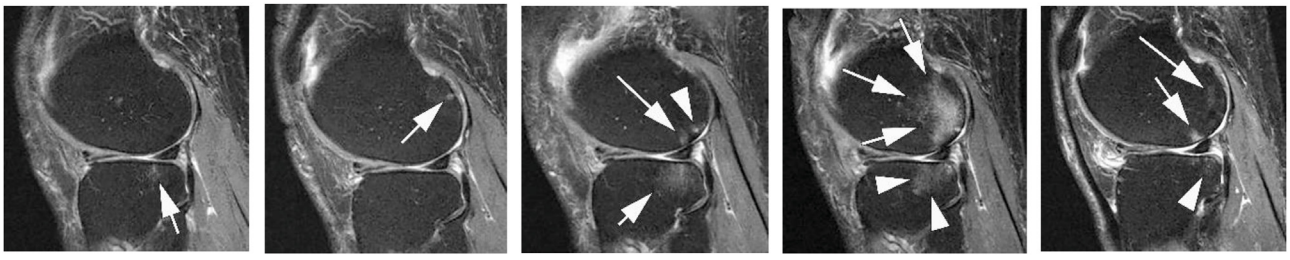


Fig. 1. Example longitudinal assessment of bone marrow lesions (BML) in the lateral tibio-femoral compartment. BMLs are fluctuating features of OA that may show decrease in some subregions and increase in others over time. A. Baseline sagittal intermediate-weighted fat-suppressed MRI shows a grade 1 MOAKS BML in the posterior lateral tibia displaying high-signal intensity, comprised of an ill-defined (edema-like) component only (arrow). B. Follow-up MRI one year later shows complete resolution of tibial BML but a dot-like small new BML at the posterior lateral femur (arrow). C. Further follow-up 1-year later shows incidence of a large grade 3 BML at the posterior lateral tibia (short arrow), complete resolution of posterior femoral BML, but new small BMLs at the posterior lateral femur (arrowhead) and central lateral femur (long arrow). D. Three years after baseline, there is a single large grade 3 BML at the posterior lateral femur that reaches into the central subregion of the lateral femur. Tibial BML show a similar size compared to previous visit (arrowheads). E. After four years, the large femoral BML has largely regressed with two remaining small grade 1 BMLs at the posterior (short arrow) and central lateral femur (long arrow). Also, the tibial BML shows decrease in size - now to ta small grade 1 lesion (arrowhead).

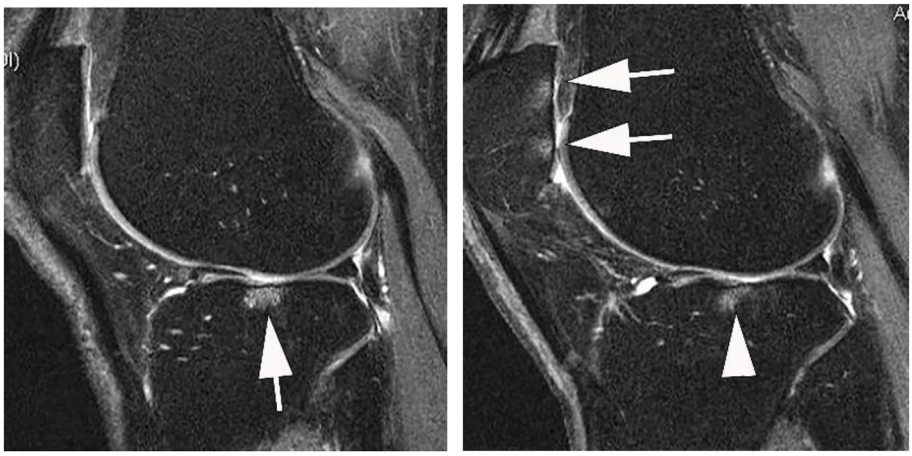


Fig. 2. Scoring of within-grade changes of structural features of OA (applied particularly to cartilage and BMLs) has shown validity and may be clinically relevant. A. Baseline sagittal intermediate weighted fat-suppressed MRI shows a grade 1 BML at the central subregion of the lateral tibia (arrow). B. At 1 year follow-up the BML is still present but has visually decreased in size (arrowhead). This will be scored as a within-grade change over time (i.e. definite visual change not fulfilling a full grade change on the MOAKS scale). In addition, there are two new BMLs at the lateral patella (arrows).

with two SRs with worsening BML score and two SRs with improving BML score would have a delta-SR score of zero. A participant with no change in BML score in any SR would also have a delta-SR score of zero. This approach should be used with caution for features that may fluctuate. It is recommended that in addition, the number of SRs with worsening and SRs with improvement are presented separately [13,16,26]. Improving scores may be expected with certain interventions, for example pharmacologic interventions with anticipated cartilage anabolic effects, and studies investigating such interventions should also consider separate assessment of worsening and improvement.

Hoffa-synovitis, Effusion-synovitis: Longitudinal changes are quantified by subtracting the follow-up score from the baseline score and can be further categorized as improvement, no change, and worsening.

Meniscus: The maximum meniscal morphology score is considered separately for the medial and lateral meniscus. The original MOAKS description does not specify how to sort the different meniscal pathologies in regard to ordinal fashion based on severity but only considers presence vs. absence of a specific tear or pathology. Past analyses have defined the following order: normal, meniscal signal, any tear type, any maceration [16,24]. Complete and progressive partial maceration is considered worse than partial maceration [14]. Furthermore and based on the available literature e.g. a complex tear will be considered "worse" (in regard to clinical impact and risk of overall structural progression) than a horizontal degenerative tear [27,28]. Modifications of MOAKS now include presence of posterior meniscal root tears that are considered exclusionary at baseline due to a markedly increased risk of rapid progressive OA [13,18]. Longitudinal change is quantified by the change in the maximum score. The number of subregions with worsening is computed by taking the number of subregions with a higher score at follow-up as compared to baseline in the anterior, body and posterior horn, separately for medial and lateral menisci. This can also be done for

Table 1
Example of Quantifying Longitudinal Change in cartilage area extent over three time points.

	LFTJ					MFTJ					PFJ			
	cLF	pLF	aLT	cLT	pLT	cMF	pMF	aMT	cMT	pMT	aLF	aMF	LP	MP
Timepoint 1	0	0	0	0	1	1	2	1	0	0	1	0	2	0
Timepoint 2	0	1	2	0	1+	1	3	1	0	0	2	0	2+	1
Timepoint 3	0	1	3	1	1+	1+	3	1	1	0	2	0	2	2

	+ denotes within-grade change (WGC) worsening		
	Maximum Score:		Number of SRs affected:
	T1: 2 T2: 3 T3: 3	Change: T1 to T2: +2 T2 to T3: +1	T1: 6 T2: 9 T3: 11	Change: T1 to T2: +3 T2 to T3: +2
	Absolute score:		Number of SRs with worsening:	
	T1: 8 T2: 14 T3: 18	Delta-Sum: T1 to T2: 6 T2 to T3: 4	exclude WGC: T1 to T2: 5 T2 to T3: 4	include WGC: T1 to T2: 7 T2 to T3: 6
	Delta-SR (exclude WGC) T1 to T2 Number of SRs with worsening: 5 Number of SRs with improvement: 0 Number of SRs with no change: 9 Delta-SR: +5 T2 to T3 Number of SRs with worsening: 4 Number of SRs with improvement: 0 Number of SRs with no change: 10 Delta-SR: +4		Delta-SR (include WGC) T1 to T2 Number of SRs with worsening: 7 Number of SRs with improvement: 0 Number of SRs with no change: 7 Delta-SR: +7 T2 to T3 Number of SRs with worsening: 6 Number of SRs with improvement: 0 Number of SRs with no change: 8 Delta-SR: +6	

Example shows results on a whole knee level. Compartmental definitions may be applied in addition.
MFTJ: medial tibiofemoral joint; LFTJ: lateral tibiofemoral joint; PFJ: patellofemoral joint; c = Central; p = Posterior; a = Anterior; LF = lateral femur; LT = lateral tibia; MF = medial femur; MT = Medial tibia; LP = lateral patella; MP = medial patella.

changes in meniscal morphology category (i.e., from normal to tear, normal to maceration or tear to maceration). Changes in meniscal extrusion are quantified as the follow-up score minus the baseline score and categorized into worsening vs. no worsening. Incident meniscal root tears are commonly considered separately due to their potential marked clinical relevance.

Examples of how various change parameters are calculated are provided in Tables 1 and 2.

3. Statistical approaches

The various approaches to quantifying longitudinal change may result in variables that are counts, ordered categories, binary categories, or continuous parameters. Each has its own set of considerations for statistical analysis. Analyzing at the SR, compartment, or whole knee level also has statistical implications. Here we describe various methods to model changes in MOAKS features as study outcomes.

Count data, such as the number of subregions with worsening MOAKS score, are often skewed and cannot be normalized with a simple transformation (e.g., natural log, square root) and thus should be modeled with a model appropriate for count outcomes [29]. Poisson regression is a generalized linear model where the outcome is modeled with a Poisson distribution, a natural log link function, and an identify function for variance (that is, mean = variance). Negative binomial regression may be needed in the case of overdispersion (excess variance). Fig. 3 shows the distribution of 24 month change in MOAKS cartilage area extent in the FNIH OA Biomarkers Consortium cohort [16]. Approximately 40% of participants experienced no change in cartilage area extent, and the count of number of subregions with worsening ranges from 0 to 8. Such data are not appropriate for linear regression.

In the case of outliers or where a dose-response relationship between predictor and outcome is not assumed, it may be advisable to categorize count data based on distribution. In the example above (Fig. 3), the number of SRs with worsening in cartilage area extent was categorized as worsening in 0, 1, 2, or 3+ SRs for analysis [16]. Categorical outcomes can be analyzed with logistic regression. Binary logistic regression is appropriate for a dichotomous outcome, e.g., any worsening vs. no worsening, and estimates the odds of outcome for those with vs. without the predictor of interest (e.g., the increased odds of worsening for participants on treatment vs. placebo). Ordinal logistic regression is appropriate for an ordered categorical outcome, e.g., worsening in 0, 1, 2, or 3+ SRs. There are various ways to model the ordering, including cumulative logits, adjacent-categories logits, and continuation ratio logits

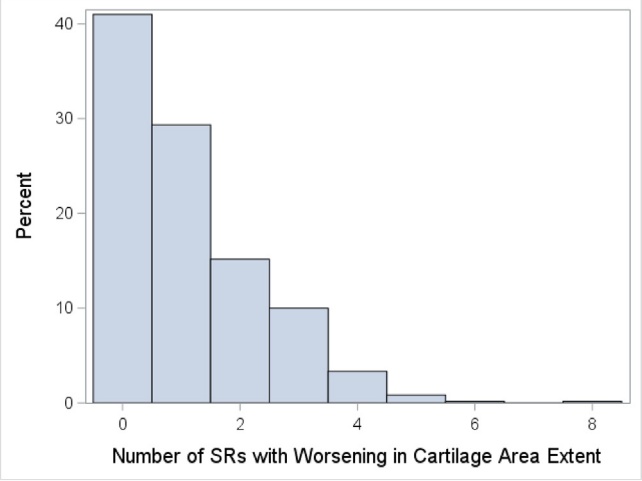


Fig. 3. The distribution of twenty-four-month change in MOAKS cartilage area extent in the FNIH OA Biomarkers Consortium cohort shows right skewness. The number of subregions with worsening in MOAKS cartilage area extent is shown on the x-axis and the percent of participants is shown along the y-axis. Over 40% of participants had zero subregions with worsening in MOAKS cartilage area extent.

Table 2
Quantifying Longitudinal Change in BMLs over three time points.

	LFTJ					MFTJ					PFJ			
	cLF	pLF	aLT	cLT	pLT	cMF	pMF	aMT	cMT	pMT	aLF	aMF	LP	MP
Timepoint 1	1	1	0	0	1	1	2	0	0	0	0	0	0	1
Timepoint 2	1-	0	0	2	1	1+	2+	0	0	0	0	2	0	0
Timepoint 3	1-	0	0	3	1+	1	2-	0	0	0	0	2	0	0

+ denotes within-grade change (WGC) worsening
- denotes within-grade change (WGC) improvement

Maximum Score:		Number of SRs affected:	
T1: 2	Change:	T1: 6	Change:
T2: 2	T1 to T2: +2	T2: 6	T1 to T2: 0
T3: 3	T2 to T3: +1	T3: 6	T2 to T3: 0
Absolute score:		Number of SRs with worsening:	
T1: 7	Delta-Sum:	exclude WGC:	include WGC:
T2: 9	T1 to T2: 2	T1 to T2: 2	T1 to T2: 4
T3: 10	T2 to T3: 1	T2 to T3: 1	T2 to T3: 2

Delta-SR (exclude WGC)	Delta-SR (include WGC)
T1 to T2	T1 to T2
Number of SRs with worsening: 2	Number of SRs with worsening: 4
Number of SRs with improvement: 2	Number of SRs with improvement: 3
Number of SRs with no change: 10	Number of SRs with no change: 7
Delta-SR: 0	Delta-SR: 1
T2 to T3	T2 to T3
Number of SRs with worsening: 1	Number of SRs with worsening: 2
Number of SRs with improvement: 0	Number of SRs with improvement: 2
Number of SRs with no change: 13	Number of SRs with no change: 10
Delta-SR: 1	Delta-SR: 0

MFTJ: medial tibiofemoral joint; LFTJ: lateral tibiofemoral joint; PFJ: patellofemoral joint; c = Central; p = Posterior; a = Anterior; LF = lateral femur; LT = lateral tibia; MF = medial femur; MT = Medial tibia; LP = lateral patella; MP = medial patella.

[30]. A detailed discussion of these models is beyond the scope of this article; each makes an assumption about the ordinal nature of the outcome. Multinomial logistic regression (generalized logits) does not assume an ordering of the outcome classes. While ignoring the ordered nature of the outcome may intuitively seem inappropriate, it allows the analyst to assess whether or which assumptions about the ordered outcome are appropriate to make about the relationship between outcome and predictor.

Continuous data, such as the delta-sum score, may have an approximate Gaussian distribution and therefore can be analyzed with methods such as linear regression. If scores are available across multiple time points, then absolute scores can be analyzed with repeated measures models, such as linear mixed effects models, and the delta-sum score can be estimated from the model. Verifying distributional assumptions (e.g., by plotting the distribution of the residuals) is essential, as delta-sum may be right skewed in some populations (e.g., early OA).

Many statistical methods, such as t-tests, Chi-square tests, and linear regression, assume that observations are independent. Special attention must be paid when analyzing repeated measures or clustered data, for example, when conducing analyses at the compartment level vs. whole knee level, or when including two vs. one knees per participant [31]. As an example, an outcome that classifies each participant as having any vs. no worsening on osteophyte score can be analyzed with a simple Chi-square test or logistic regression model. If instead each participant contributes two knees, and each knee is categorized as having any vs. no worsening in osteophyte score (two observations per participant), or if any vs. no osteophyte worsening is computed at the compartment level (LFTJ, MFTJ, PFJ: three observations per knee) then models that account for the correlated nature of the data (knees nested within participants, compartments nested within knees) are required. There are many methods available for clustered data, such as random-effect, mixed-effect, and marginal models; a thorough review of such methods is presented elsewhere [32,33].

MOAKS includes scores across multiple joint features, compartments, and SRs, allowing for the computation of potentially many outcomes. With nested data it may or may not be of interest to test for differences in level; e.g., in a model with two knees per person it may not be of interest to test whether outcomes are different in left vs. right knee, but may be of interest to test for differences in injured vs. uninjured knee. In the case of nested models with compartments or SRs, pairwise comparisons should be approached cautiously. Investigators using MOAKS must be aware of the problem of multiple testing: testing multiple hypotheses without correction to the alpha level can result in inflated Type I error rates. Many methods exist to correct alpha for multiple comparisons [34].

4. Examples

Roemer and colleagues used the delta-sum and delta-SR approach in a post-hoc analysis of the effects of sprifermin on cartilage and non-cartilaginous joint tissues [35]. Mann–Whitney − Wilcoxon tests assessed differences between treatment groups. The authors found significant differences between the treatment groups in cartilage area extent in the PFJ (delta-sum approach; treatment 0.05, 95 % CI (−0.06, 0.17) vs. placebo 0.44, 95 % CI (−0.18, 1.06), $p = 0.048$; delta-SR approach; treatment (0.02 95 % CI (−0.04, 0.08) vs. placebo 0.22, 95 % CI (−0.05, 0.49), $p = 0.046$). While no significant differences were found between treatment groups for BMLs, this paper illustrates how the two approaches quantify longitudinal change for parameters that may improve over time: the delta-sum and delta-SR values for the treatment group were both negative, indicating overall improvement in BML score.

Collins and colleagues used the delta-sum approach to quantify 5-year structural changes in the MeTeOR (Meniscal Tear in Osteoarthritis Research) trial, a multicenter RCT of arthroscopic partial meniscectomy vs. physical therapy in patients with degenerative meniscal tear and mild-to-moderate knee OA [36]. The authors calculated the absolute score separately for cartilage area extent, full-thickness loss, osteophytes, and BMLs at baseline, 18-months, and 5-years follow-up and then used linear mixed-effects models to assess the association between treatment group and score and to estimate changes in score, i.e., delta-sum. The authors found significant differences in delta-sum osteophyte score between treatment groups over both the baseline to 18 month and 18–60-month time periods, with the APM group demonstrating higher delta-sum scores (i.e., more worsening).

The RESTORE RCT evaluated the effects of intra-articular platelet rich plasma (PRP) injections on symptoms and joint structure in patients with knee OA [37]. Secondary MRI outcomes included the number of subregions with worsening in full-thickness cartilage damage (categorized as 0, 1, 2, or $\geq$3), change in whole knee effusion-synovitis (categorized as worsened, no change, or improved), any worsening in Hoffa-synovitis (referred to as intercondylar in this study), and any worsening in BML size across all subregions. Changes in full-thickness cartilage damage and whole knee effusion-synovitis were analyzed with multinomial logistic regression and worsening in Hoffa-synovitis and BML size were analyzed with log-binomial regression models. Seventeen percent of the PRP group demonstrated worsening in full-thickness cartilage damage in 3 or more subregions compared to 7% of the placebo group for an increased odds of worsening of 2.7 (95% CI: 1.16–6.34).

Roemer et al. and Collins et al. demonstrate how change in maximum grade and change in number of SRs can be used to quantify longitudinal change in the FNIH OA Biomarkers study, and how within-grade changes can be included or excluded from these calculations [16,26]. The FNIH OA Biomarkers study was a nested case-control study that aimed to evaluate the longitudinal validity of imaging and biochemical markers as efficacy of intervention biomarkers in knee OA. Case knees were those with joint space narrowing and pain worsening over 48 months. 24-month changes in cartilage area extent were associated with case status in analyses both including and excluding within-grade change; 68% of the control group and 54% of the case group had no worsening in cartilage area extent when within-grade changes were excluded; when within-grade changes were excluded the percent with no change dropped to 48% and 27% respectively. Additional details of the cartilage results of the FNIH study are presented in Fig. 3.

5. Discussion

In this perspective we described how whole-organ SQ scoring of MRI i.e., MOAKS can be used to quantify longitudinal change in knee OA. The details captured in MOAKS or comparable systems like WORMS provide a nuanced picture of disease progression across multiple tissues. While SQ scoring has been proposed for disease phenotyping and eligibility assessment in clinical trials of knee OA [13,18], the outcomes described here may also be used as trial outcomes to assess structural disease progression and evaluate treatment efficacy.

Here we present recommendations for summarizing disease progression within specific tissue types. There has been interest in whether and how a composite outcome reflecting disease progression across the entire joint might be defined. The OMERACT Technical Advisory Group defines a multi-outcome domain (MOD) as “a within-patient combination of component outcomes, and an individual patient's evaluation depends on the observation of all of the components in that patient with a single overall rating determined according to a specified rule” and a composite outcome domain (COD) as “a number of component outcomes and is defined as the occurrence in a patient of one, some or all of these specified components” [38]. A composite outcome in a DMOAD trial might be reaching either a pre-specified pain threshold or undergoing total knee replacement [39]. Driban and colleagues attempted to define a multi-outcome domain using quantitative measures of cartilage damage, BMLs, and effusion-synovitis volume to reflect damage across multiple tissues [40]. The authors report two composite scores: a ‘cumulative cartilage damage’ score reflecting structural progression and a ‘BML plus effusion-synovitis’ score reflecting symptomatic progression. SQ scoring systems such as MOAKS, which provide rich detail across multiple tissues, may provide an opportunity to further investigate multi-outcome

domains for DMOAD trials. While it is unclear whether and how features should be combined across domains, unsupervised machine learning methods could help inform how and when to combine features [41].

Patterns of both patient-reported symptoms and of structural damage vary substantially between knee OA patients, suggesting that different phenotypes (subpopulations) of OA exist [42,43]. As such, a "one-size fits all" approach is unlikely to work in developing therapies for knee OA – interventions must be tailored to and tested in the appropriate population [44]. The Rapid OsteoArthritis MRI Eligibility Score (ROAMES) provides a framework for how SQ scoring can be used to perform phenotypic stratification of DMOAD clinical trial participants [18]. The multi-tissue assessment provided by SQ scoring systems such as MOAKS provides an opportunity to not only stratify at trial enrollment but also to match phenotype to outcome. A challenge to be resolved in future work is based on the fact that structural phenotype may overlap and one joint may exhibit more than one phenotype.

A DMOAD trial outcome must be sensitive to change [45]. There are few studies directly comparing the sensitivity to change of quantitative vs. SQ MRI assessments. Some work has suggested that quantitative measures of cartilage thickness may be more sensitive to change than SQ measures, while others have found sensitivity to change to be similar [45,46]. Including within-grade changes in SQ MRI scores has been shown to improve sensitivity to change for both cartilage and BML assessment [22]. Recent studies have shown that both full-grade and within-grade changes in MOAKS cartilage scores correspond with longitudinal quantitative cartilage thickness loss [23,47]. While standardized response means (SRMs) are often used to assess responsiveness, the interpretation of this measure (mean change divided by standard deviation of change) is less intuitive for non-Gaussian outcomes, making direct comparison between SQ and quantitative measures difficult. Recent findings from RCTs may shed light on sensitivity to change. For example, the RESTORE RCT, assessing the effect of intra-articular platelet-rich plasma vs placebo injection, found no significant difference in structural endpoints between treatment groups and. The study finds that approximately 50% of participants experienced areas with cartilage thinning and 15% had worsening of full-thickness cartilage damage over 18 months [37]. The SRM for the primary outcome of quantitative percentage change in medial tibial cartilage volume was approximately 0.2. Over 18 months in the MeTeOR trial approximately 60% of non-operatively treated participants experienced worsening in cartilage surface area extent score and 50% in full-thickness cartilage loss score [48].

To-date, no MRI-based measures of disease progression are approved by FDA as surrogate endpoints for DMOAD trials. FDA guidance does not refer to the type of surrogate endpoint (e.g., SQ vs. quantitative measure, imaging vs. biochemical biomarker), but does require that there be "substantial confidence, either based on empirical evidence from randomized, controlled comparisons from clinical trials and/or based on a comprehensive understanding of the disease process and product mechanism of action, that an effect on the candidate structural endpoint will reliably predict an effect on the clinical outcomes of interest." [49] There is precedent for the use of SQ scoring of imaging as trial outcomes, such as the Sharp or modified-Sharp score for erosion and joint space narrowing in rheumatoid arthritis [50]. Future work should investigate establishing a minimal clinically important difference (MCID) for OA structural endpoints.

Another major hurdle in assessing treatment efficacy in DMOAD trials is the lack of validated structural endpoints and the uncertainty around the magnitude of change in these endpoints that would translate to a clinically meaningful benefit to patients [6,49]. The rich data provided by SQ assessment provides an opportunity to investigate which combination of features best measure how a patient feels or functions, or how the joint survives. However, determining when and how to summarize these data does not come without challenges. For example, while summarizing scores across all subregions results in a single continuous outcome measure and facilitates the use of traditional analytic approaches, it leads to a loss of information, as sums are the result of a spectrum of affected subregions and severity grades. A high total score (i.e., summing across subregions) could reflect small amounts of damage in many subregions, or large amounts of damage in few subregions – a fact that is also relevant for defining longitudinal change. It may also obscure changes for features with fluctuating damage (e.g., BMLs) and thus lead to challenges in correlating changes in structural features with symptoms. Prior work by Runhaar and colleagues attempted to define subregional osteoarthritis progression using MOAKS [21]. The authors define thresholds for both improvement and progression in BMLs, cartilage defects, osteophytes, and meniscal pathologies. While this is an important step in conceptualizing SQ-based progression, the guidance we present here represents a more general approach to defining change that could be tailored to a specific trial or intervention.

To date no disease-modifying therapies are approved for knee OA. SQ imaging assessment of longitudinal change provides an opportunity to better understand disease progression across multiple tissue types, which may allow future trials to match outcome to patient phenotype or to treatment mechanism of action. The detailed data provided by these assessments will allow researchers to determine whether and how to best combine scores such that changes are reflective of clinically relevant outcomes and to determine what magnitude of change represents clinical benefit.

Author contributions

Substantial contributions to the conception or design of the work: all authors.

Drafting the work or reviewing it critically for important intellectual content: all authors.

Final approval of the version to be published: all authors.

Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: all authors.

Conflicts of interest

JEC has received consultancies from BICL, LLC.

FWR is Chief Medical Officer and shareholder of BICL, LLC. and has received consultancies, speaking fees, and/or honoraria from Calibr –California Institute of Biomedical Research and Grünenthal, GmbH.

AG has received consultancies, speaking fees, and/or honoraria from Pfizer, Novartis, AstraZeneca, Merck Serono, and TissueGene and is President and shareholder of Boston Imaging Core Lab (BICL), LLC a company providing image assessment services.

Acknowledgements

We report no other contributors or funding sources to acknowledge.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2024.100465.

References

[1] T. Vos, A.A. Abajobir, K.H. Abate, C. Abbafati, K.M. Abbas, F. Abd-Allah, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet 390 (2017) 1211–1259.

[2] B.R. Deshpande, J.N. Katz, D.H. Solomon, E.H. Yelin, D.J. Hunter, S.P. Messier, et al., Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity, Arthritis Care Res. 68 (2016) 1743–1750.

[3] M. Cross, E. Smith, D. Hoy, S. Nolte, I. Ackerman, M. Fransen, et al., The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study, Ann. Rheum. Dis. 73 (2014) 1323–1330.

[4] R.R. Bannuru, C.H. Schmid, D.M. Kent, E.E. Vaysbrot, J.B. Wong, T.E. McAlindon, Comparative effectiveness of pharmacologic interventions for knee osteoarthritis: a systematic review and network meta-analysis, Ann. Intern. Med. 162 (2015) 46–54.

[5] W. Yang, C. Sun, S.Q. He, J.Y. Chen, Y. Wang, Q. Zhuo, The efficacy and safety of disease-modifying osteoarthritis drugs for knee and hip osteoarthritis-a systematic review and network meta-analysis, J. Gen. Intern. Med. 36 (2021) 2085–2093.
[6] J.S. Kim, S. Borges, D.J. Clauw, P.G. Conaghan, D.T. Felson, T.R. Fleming, et al., FDA/Arthritis Foundation osteoarthritis drug development workshop recap: assessment of long-term benefit, in: Seminars in Arthritis and Rheumatism, Elsevier, 2022 152070.
[7] D. Hayashi, F.W. Roemer, A. Guermazi, How to effectively utilize imaging in disease-modifying treatments for osteoarthritis clinical trials: the radiologist's perspective, Expert Rev. Mol. Diagn. 21 (2021) 673–684.
[8] F.W. Roemer, D.T. Felson, J.J. Stefanik, G. Rabasa, N. Wang, M.D. Crema, et al., Heterogeneity of cartilage damage in Kellgren and Lawrence grade 2 and 3 knees: the MOST study, Osteoarthritis Cartilage 30 (2022) 714–723.
[9] A. Guermazi, F.W. Roemer, D. Burstein, D. Hayashi, Why radiography should no longer be considered a surrogate outcome measure for longitudinal assessment of cartilage in knee osteoarthritis, Arthritis Res. Ther. 13 (2011) 247.
[10] S.A. Mazzuca, K.D. Brandt, P.A. Dieppe, M. Doherty, B.P. Katz, K.A. Lane, Effect of alignment of the medial tibial plateau and x-ray beam on apparent progression of osteoarthritis in the standing anteroposterior knee radiograph, Arthritis Rheum. 44 (2001) 1786–1794.
[11] F.W. Roemer, C.K. Kwoh, M.J. Hannon, D.J. Hunter, F. Eckstein, T. Fujii, et al., What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging–based trajectory analysis over four years in the Osteoarthritis Initiative, Arthritis Rheumatol. 67 (2015) 2085–2096.
[12] F.W. Roemer, D.J. Hunter, M.D. Crema, C.K. Kwoh, E. Ochoa-Albiztegui, A. Guermazi, An illustrative overview of semi-quantitative MRI scoring of knee osteoarthritis: lessons learned from longitudinal observational studies, Osteoarthritis Cartilage 24 (2016) 274–289.
[13] A. Guermazi, F.W. Roemer, M.D. Crema, M. Jarraya, A. Mobasheri, D. Hayashi, Strategic application of imaging in DMOAD clinical trials: focus on eligibility, drug delivery, and semiquantitative assessment of structural progression, Ther Adv Musculoskelet Dis 15 (2023), 1759720X231165558.
[14] D.J. Hunter, A. Guermazi, G.H. Lo, A.J. Grainger, P.G. Conaghan, R.M. Boudreau, et al., Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score), Osteoarthritis Cartilage 19 (2011) 990–1002.
[15] D.J. Hunter, W. Zhang, P.G. Conaghan, K. Hirko, L. Menashe, L. Li, et al., Systematic review of the concurrent and predictive validity of MRI biomarkers in OA, Osteoarthritis Cartilage 19 (2011) 557–588.
[16] J.E. Collins, E. Losina, M.C. Nevitt, F.W. Roemer, A. Guermazi, J.A. Lynch, et al., Semiquantitative imaging biomarkers of knee osteoarthritis progression: data from the foundation for the national institutes of health osteoarthritis biomarkers consortium, Arthritis Rheumatol. 68 (2016) 2422–2431.
[17] C. Peterfy, A. Guermazi, S. Zaim, P. Tirman, Y. Miaux, D. White, et al., Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis, Osteoarthritis Cartilage 12 (2004) 177–190.
[18] F.W. Roemer, J. Collins, C.K. Kwoh, M.J. Hannon, T. Neogi, D.T. Felson, et al., MRI-based screening for structural definition of eligibility in clinical DMOAD trials: rapid OsteoArthritis MRI Eligibility Score (ROAMES), Osteoarthritis Cartilage 28 (2020) 71–81.
[19] W. Wirth, S. Maschek, A. Wisser, A. Guermazi, D.J. Hunter, C.K. Kwoh, et al., Selection of knees with subsequent cartilage thickness loss based on magnetic resonance imaging semiquantitative grading: data from the osteoarthritis initiative foundation for the national institutes of health biomarker cohort, Arthritis Care Res 75 (8) (2023 Aug) 1773–1782.
[20] F.W. Roemer, C.K. Kwoh, D. Hayashi, D.T. Felson, A. Guermazi, The role of radiography and MRI for eligibility assessment in DMOAD trials of knee OA, Nat. Rev. Rheumatol. 14 (2018) 372–380.
[21] J. Runhaar, D. Schiphof, B. van Meer, M. Reijman, S.M. Bierma-Zeinstra, E.H. Oei, How to define subregional osteoarthritis progression using semi-quantitative MRI osteoarthritis knee score (MOAKS), Osteoarthritis Cartilage 22 (2014) 1533–1536.
[22] F.W. Roemer, M.C. Nevitt, D.T. Felson, J. Niu, J.A. Lynch, M.D. Crema, 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 20 (2012) 1391–1398.
[23] F. Roemer, S. Maschek, A. Wisser, A. Guermazi, D. Hunter, F. Eckstein, et al., Worsening of articular tissue damage as defined by semi-quantitative MRI is associated with concurrent quantitative cartilage loss over 24 months, Cartilage 14 (2023) 39–47.
[24] F.W. Roemer, M. Jansen, A.C.A. Marijnissen, A. Guermazi, R. Heiss, S. Maschek, et al., Structural tissue damage and 24-month progression of semi-quantitative MRI biomarkers of knee osteoarthritis in the IMI-APPROACH cohort, BMC Muscoskel. Disord. 23 (2022) 988.
[25] F.W. Roemer, J. Kraines, A. Aydemir, S. Wax, M.C. Hochberg, M.D. Crema, et al., Evaluating the structural effects of intra-articular sprifermin on cartilage and non-cartilaginous tissue alterations, based on sqMRI assessment over 2 years, Osteoarthritis Cartilage 28 (2020) 1229–1234.
[26] F.W. Roemer, A. Guermazi, J.E. Collins, E. Losina, M.C. Nevitt, J.A. Lynch, et al., Semi-quantitative MRI biomarkers of knee osteoarthritis progression in the FNIH biomarkers consortium cohort - methodologic aspects and definition of change, BMC Muscoskel. Disord. 17 (2016) 466.
[27] J.B. Driban, J.E. Davis, B. Lu, L.L. Price, R.J. Ward, J.W. MacKay, et al., Accelerated knee osteoarthritis is characterized by destabilizing meniscal tears and preradiographic structural disease burden, Arthritis Rheumatol. 71 (2019) 1089–1100.
[28] J.T. Badlani, C. Borrero, S. Golla, C.D. Harner, J.J. Irrgang, The effects of meniscus injury on the development of knee osteoarthritis: data from the osteoarthritis initiative, Am. J. Sports Med. 41 (2013) 1238–1244.
[29] A.C. Cameron, P.K. Trivedi, Regression Analysis of Count Data, ume 53, Cambridge University Press, 2013.
[30] A.J. Dobson, A.G. Barnett, An Introduction to Generalized Linear Models, CRC Press, 2018.
[31] J. Ranstam, Repeated measurements, bilateral observations and pseudoreplicates, why does it matter? Osteoarthritis Cartilage 20 (2012) 473–475.
[32] Y.-H.H. Chang, M.R. Buras, J.M. Davis, C.S. Crowson, Avoiding blunders when analyzing correlated data, clustered data, or repeated measures, J. Rheumatol. 50 (2023) 1269–1272.
[33] L.M. Sullivan, K.A. Dukes, E. Losina, An introduction to hierarchical linear modelling, Stat. Med. 18 (1999) 855–888.
[34] J. Ranstam, Multiple P-values and bonferroni correction, Osteoarthritis Cartilage 24 (2016) 763–764.
[35] F.W. Roemer, A. Aydemir, S. Lohmander, M.D. Crema, M.D. Marra, N. Muurahainen, et al., Structural effects of sprifermin in knee osteoarthritis: a post-hoc analysis on cartilage and non-cartilaginous tissue alterations in a randomized controlled trial, BMC Muscoskel. Disord. 17 (2016) 267.
[36] J.E. Collins, S. Shrestha, E. Losina, R.G. Marx, A. Guermazi, M. Jarraya, et al., Five-year structural changes in the knee among patients with meniscal tear and osteoarthritis: data from a randomized controlled trial of arthroscopic partial meniscectomy versus physical therapy, Arthritis Rheumatol. 74 (2022) 1333–1342.
[37] K.L. Bennell, K.L. Paterson, B.R. Metcalf, V. Duong, J. Eyles, J. Kasza, et al., Effect of intra-articular platelet-rich plasma vs placebo injection on pain and medial tibial cartilage volume in patients with knee osteoarthritis: the RESTORE randomized clinical trial, JAMA 326 (2021) 2021–2030.
[38] G.A. Wells, P. Tugwell, G. Tomasson, F. Guillemin, L.J. Maxwell, B.J. Shea, et al., Composite outcomes at OMERACT: multi-outcome domains and composite outcome domains, Semin. Arthritis Rheum. 51 (2021) 1370–1377.
[39] Y. Kim, G. Levin, N.P. Nikolov, R. Abugov, R. Rothwell, Concept end points informing design considerations for confirmatory clinical trials in osteoarthritis, Arthritis Care Res. 74 (2022) 1154–1162.
[40] J.B. Driban, L.L. Price, M.P. LaValley, G.H. Lo, M. Zhang, M.S. Harkey, et al., Novel framework for measuring whole knee osteoarthritis progression using magnetic resonance imaging, Arthritis Care Res. 74 (2022) 799–808.
[41] J. Niu, D.T. Felson, T. Neogi, M.C. Nevitt, A. Guermazi, F. Roemer, et al., Patterns of coexisting lesions detected on magnetic resonance imaging and relationship to incident knee osteoarthritis: the multicenter osteoarthritis study, Arthritis Rheumatol. 67 (2015) 3158–3165.
[42] N.E. Lane, K. Brandt, G. Hawker, E. Peeva, E. Schreyer, W. Tsuji, et al., OARSI-FDA initiative: defining the disease state of osteoarthritis, Osteoarthritis Cartilage 19 (2011) 478–482.
[43] M.A. Karsdal, M. Michaelis, C. Ladel, A.S. Siebuhr, A.R. Bihlet, J.R. Andersen, et al., Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future, Osteoarthritis Cartilage 24 (2016) 2013–2021.
[44] S.M. Bierma-Zeinstra, M. van Middelkoop, Osteoarthritis: in search of phenotypes, Nat. Rev. Rheumatol. 13 (2017) 705–706.
[45] F. Eckstein, A. Guermazi, G. Gold, J. Duryea, M.-P.H. Le Graverand, W. Wirth, et al., Imaging of cartilage and bone: promises and pitfalls in clinical trials of osteoarthritis, Osteoarthritis Cartilage 22 (2014) 1516–1532.
[46] D. Hunter, W. Zhang, P. Conaghan, K. Hirko, L. Menashe, W. Reichmann, et al., Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence, Osteoarthritis Cartilage 19 (2011) 589–605.
[47] F. Roemer, S. Maschek, A. Wisser, A. Guermazi, D. Hunter, W. Wirth, Relevance of within-grade cartilage assessment using the moaks scoring instrument and corresponding quantitative cartilage thickness loss: data from the OAI FNIH Biomarkers Consortium Study, Osteoarthritis Cartilage 30 (2022) S267–S268.
[48] J.E. Collins, E. Losina, R.G. Marx, A. Guermazi, M. Jarraya, M.H. Jones, et al., Early magnetic resonance imaging-based changes in patients with meniscal tear and osteoarthritis: eighteen-month data from a randomized controlled trial of arthroscopic partial meniscectomy versus physical therapy, Arthritis Care Res. 72 (2020) 630–640.
[49] Guidance for industry: osteoarthritis. Structural endpoints for the development of drugs, devices, and biological products for treatment, in: Rockville (MD): US Food and Drug Administration, 2018.
[50] Y.-J. Park, A.M. Gherghe, D. van der Heijde, Radiographic progression in clinical trials in rheumatoid arthritis: a systemic literature review of trials performed by industry, RMD Open 6 (2020) e001277.