Quantitative three-dimensional assessment of knee joint space width from weight-bearing CT

Journal:	*Radiology*
Manuscript ID	RAD-20-3928.R1
Manuscript Type:	Technical Development
Manuscript Categorization Terms:	CT < 2. MODALITIES/TECHNIQUES, CT-Quantitative < 2. MODALITIES/TECHNIQUES, Image Postprocessing < 2. MODALITIES/TECHNIQUES, Skeletal-Appendicular < 4. AREAS/SYSTEMS, Knee < 5. STRUCTURES, Joints < 5. STRUCTURES, Arthritides < 6. TOPICS, Computer Applications-3D < 6. TOPICS, Segmentation < 6. TOPICS, Statistics < 7. METHODOLOGY, Observer Performance < 7. METHODOLOGY, Experimental Investigations < 7. METHODOLOGY

TITLE

Quantitative three-dimensional assessment of knee joint space width from weight-bearing CT

ARTICLE TYPE

Technical development

SUMMARY

Joint space mapping of weight-bearing knee CT can deliver personalized quantitative measurements of joint space width in three dimensions that are structurally relevant in osteoarthritis, learnable by novice users, and highly repeatable.

KEY RESULTS

1. A three-dimensional surface-based approach to measurement, display, and analysis of joint space width can be delivered from weight-bearing knee CT imaging.
2. The joint space mapping technique is highly reproducible and can demonstrate smallest detectable differences of less than ±0.1mm.
3. Threshold differences in joint space width can be clearly demonstrated for an individual compared to prior imaging for patient-specific interpretation.

ABBREVIATIONS

3-D = three-dimensions/three-dimensional; JSM = joint space mapping; JSW = joint space width; KLG = Kellgren and Lawrence grade; LOA = limits of agreement; RMSCV = root mean square coefficient of variation.

ABSTRACT

Background:

Imaging of structural disease in osteoarthritis has traditionally relied on MRI and radiography. Joint space mapping (JSM) can quantitatively map joint space width (JSW) in three dimensions (3-D) from CT.

Purpose:

To demonstrate reproducibility, repeatability, and feasibility of JSM at the knee using weight-bearing CT.

Materials and Methods:

Two convenience samples of weight-bearing CT of both knees acquired from 2014 to 2018 and radiographic Kellgren and Lawrence grade (KLG) ≤2 were analyzed retrospectively with JSM to deliver 3-D JSW maps. For reproducibility, three sets of knees were used for novice training, then JSM output was compared against an expert. JSM was also performed on 2-week follow-up imaging in the second cohort yielding 3-D JSW difference maps for repeatability. Statistical parametric mapping was performed on all knees (KLG=0-4) to show feasibility of surface-based analysis in 3-D.

Results:

Reproducibility (20 individuals, 58±7 years, body mass index 28±6 kg/m^2, 14 women) and repeatability (9 individuals, 53±6 years, 26±4 kg/m^2, 7 women) reached best performance of less than ±0.1mm in the central medial tibiofemoral joint space for individuals without radiographic disease. Average root mean square coefficient of variation values were <5% across all groups. Statistical parametric mapping (33 individuals, 57±7 years, 27±6 kg/m^2, 23 women) showed that the central-to-posterior medial joint space was significantly narrower by 0.5 mm for each increment in KLG (threshold $p<.05$). A single knee (KLG=2) demonstrated baseline versus 24-month change in 3-D JSW distribution beyond smallest detectable difference across the lateral joint space.

Conclusion:

Joint space mapping is feasible at the knee with weight-bearing CT, demonstrating a relationship between three-dimensional joint space width distribution and structural joint disease. It is reliably learned by novice users, can be personalized to disease phenotypes, and can achieve a smallest detectable difference at least 50% better than the reported best performance of radiography.

INTRODUCTION

Substantial challenges remain in the search for disease-modifying treatments against osteoarthritis but improving sensitivity of imaging remains an important goal, particularly for identifying early disease, monitoring progression, and stratifying patients for therapeutic options. Up until recently, radiographic minimum joint space width (JSW) has been the mainstay of imaging assessment in clinical trials as approved by the U.S. Food and Drug Administration, however there is now recognition of the value that cross-sectional imaging could bring in this role(1).

Quantitative imaging of osteoarthritis has usually been considered the reserve of physiological MRI such as delayed gadolinium-enhanced MRI of cartilage, or dGEMRIC, T2, T2*, and T1rho recovery/relaxation mapping(2), while morphologic MRI has focused on cartilage thickness usually reduced to a single scalar value by subregion(3). Semi-quantitative systems such as the MRI Osteoarthritis Knee Score(4), radiographic atlases(5), and the Kellgren and Lawrence score(6) are often used for grading disease features, while minimum JSW remains the favored quantitative radiographic measure(7).

Although CT is established in the role of visualizing mineralized peri-articular structures such as subchondral bone(8), it has otherwise been limited when compared to radiography in the investigation of the joint space in osteoarthritis by an historic inability to image in a weight-bearing position. However, cone beam CT technology can now acquire images of the standing knee(9), meaning that there could be advantages over radiography in the assessment of JSW in a three-dimensions (3-D), principally through removal of projectional variability in X-ray beam positioning, and the opportunity for a more accurate representation of JSW distribution leading to greater sensitivity in detecting disease-relevant structural changes. Although this may come with some increase in image noise, it can be achieved at much lower doses than, for example, clinical hip CT (~0.02-0.25 mSv compared to ~2.5 mSv per scan)(10–12).

Here we apply a technique called joint space mapping (JSM) to measure, display, and statistically analyze knee joint space width in 3-D from weight-bearing CT. We show the implementation of JSM at the knee, inter-operator reproducibility, test-retest repeatability, and feasibility of 3-D surface-based statistical analysis, then demonstrate its application and discuss how to establish clinical utility of the technique. Reproducibility and repeatability is tested in the context of individuals with no or early radiographic osteoarthritis as the population most likely to benefit from disease-modifying intervention.

MATERIALS AND METHODS

Joint space mapping of the knee

Steps from bone segmentation through to 3-D JSW map creation are shown in Figure 1, all performed using free-to-download in-house StradView software (currently v6.13, https://mi.eng.cam.ac.uk/Main/StradView, Graham Treece, Cambridge University Engineering Department, Cambridge, U.K.). This takes <10 minutes for each knee, most of this time spent on initial bone segmentation. A final JSW map consists of 500-1000 independently measured data points at each joint. Each map shows the distance between the femoral and tibial subchondral bone surfaces (Figure 2), thus also defining the location of each joint bone surface in 3-D. We use the halfway distance between these two subchondral bone surfaces as a skeleton of the whole joint (Figure 3). Since this halfway surface is different for every knee, a 'canonical' halfway surface was created as the right-sided average from all knees in these studies to provide a common frame of reference on which results are subsequently displayed and analyzed.

The canonical surface is automatically registered to each individual patch with a similarity transform (manipulation of scale, positioning, and mirroring), followed by a locally affine deformation (Figure E1 [online]) using free-to-download in-house software called wxRegSurf (currently v20, http://mi.eng.cam.ac.uk/~ahg/wxRegSurf/, Andrew Gee, Cambridge University Engineering Department, Cambridge, U.K.). These registrations use an iterative closest point algorithm that also allows matching of the patch perimeter. wxRegSurf automatically transfers JSW measurements from each vertex to the nearest neighbor on the canonical surface, with some blurring of data before and after transfer to prepare for subsequent statistical parametric mapping.

Reproducibility

We used a convenience sample of weight-bearing CT images of both knees from 23 participants in the Multicenter Osteoarthritis Study (MOST) from 2016 to 2018 (Figure 4) that were involved in a prior study comparing standing CT with radiographic JSW(7). Participants were recruited to MOST with University of Iowa Institutional Review Board approval 20003064 for demographic data collection and 201602741 for weight-bearing CT acquisition (all under FWA00003007) with informed consent prior to enrolment.

A prototype commercial CT scanner (LineUp, CurveBeam LLC, Hatfield, PA, USA) took weight-bearing images of both knees in a fixed-flexion stance using a SynaFlexer plexiglass positioning frame

(BioClinica (formerly Synarc), San Francisco, CA, U.S.A.). Data sets with isotropic voxels at 0.37 mm and an axial 200 x 350 mm field of view were reconstructed from cone beam projections. Typical effective dose for each examination was 0.024 mSv with a CTDIvol of 1.1 mGy and DLP of 22 mGy.cm.

All distal femurs were segmented by a single novice operator (SL) trained by an expert in bone segmentation with 10 years' experience (TT); both were blinded seeing only the imaging data and anonymized study ID. Training of a novice with no prior experience in joint space patch segmentation (SL) consisted of working on three bilateral knee sets selected out with KLGs 0-3 under the guidance of an expert with 8 years' experience (TT). Joint space perimeters were segmented blindly and independently by both for the remaining 38 knees with KLG0=25, KLG1=7, and KLG2=6. Knees with KLG=3&4 were excluded because they represented advanced structural disease. JSM was performed at each knee with JSW data transferred to, presented, and analyzed on the canonical halfway joint surface. Bland-Altman analysis for bias and limits of agreement with 95% confidence intervals were calculated across this surface with a method that controlled for intra-subject correspondence if two knees in a study group were from the same individual(13).

Test-retest repeatability

We used a convenience sample of bilateral weight-bearing knee CT imaging from 30 individuals recruited between June and August 2014 for another study looking at the test-retest repeatability of a different JSW measurement methodology(11). Individuals with motion artefact and KLG3-4 were excluded (Figure 4) leaving 14 knees from 9 participants with KLG0=4, KLG1=3, and KLG2=7. University of Iowa Institutional Review Board approval 201403723 had been obtained with informed oral consent from all participants for measurement of 3-D JSW distribution. Each participant agreed to a baseline and repeat scan at 2 weeks.

The same prototype CT scanner from CurveBeam as for the reproducibility study was used with the same dosage metrics and imaging data reconstruction as above. Participants were imaged with knees in a 20° fixed-flexion position, but without SynaFlexer frame support.

All femurs were segmented by a single novice operator (SR) trained by an expert in bone segmentation with 10 years' experience (TT), again both blinded to all but the imaging data and anonymized study ID. JSM was performed on baseline knees by TT. The distal femur segmented at baseline was rigidly registered (i.e., translation and rotation only) to the same side femur segmented

at follow-up using wxRegSurf to align them between attendances. The same rigid transformation was applied to the baseline joint space patch, with the rest of the JSM process applied to the baseline patch in the follow-up data volume. After registering the canonical halfway surface to each individual halfway surface produced by JSM, results from visit 2 were subtracted from visit 1 and repeatability statistics calculated with the same Bland-Altman method as for the reproducibility analysis(13).

Feasibility of joint space mapping

We used a combined convenience sample of weight-bearing knee CT from the groups above (Figure 4) including all KLGs (KLG0=31, KLG1=12, KLG2=14, KLG3=7, and KLG4=2) from the 33 participants analyzed with JSM by a single user (TT). According to principal component analysis of canonical surface registration vectors to each individual surface, 90% of total shape variation was accounted for by the first five modes.

Statistical parametric mapping is an established technique in functional neuroimaging that uses a general linear model at each point on a surface to account for variability in 3-D surface data in terms of experimental and confounding factors(14). For this feasibility study, in the absence of outcome data we used an experimental term of KLG as a measure of structural disease and confounding terms of age, sex, body mass index, and the first five shape modes to control for any effects of systematic misregistration(15), with a threshold p value of .05. Statistical parametric mapping analysis was performed using the freely downloadable SurfStat package (https://www.math.mcgill.ca/keith/surfstat/, developed by Keith Worsley, Department of Mathematics and McGill University, Québec, Canada) in MATLAB R2109b (© 1984–2019, The MathWorks, Inc., Natick, MA, U.S.A.).

Personalized joint space mapping

It is also possible to visualize changes in 3-D JSW for an individual by comparing baseline and follow-up imaging on the canonical joint surface. Here we use the example of baseline and 24-month follow up imaging for an individual with KLG=2. IRB approval and consent for use of this data was the same as for the reproducibility study. Histogram distribution, median, and interquartile range values were calculated, also showing in 3-D where follow-up JSW has reduced beyond the baseline values. By taking the KLG=2 limits of agreement map as a mask, we also show where recorded differences in JSW are within the smallest detectable difference of the JSM technique for the KLG=2 category.

RESULTS

Reproducibility of joint space mapping

The reproducibility study included 20 individuals, mean age ± SD 58 ± 7 years, body mass index 28 ± 6 kg/m^2, 14 women (Table 1). Results are summarized as values averaged across the whole joint surface and broken down by subcategories of KLG<2 and KLG=2 (Table 2). Looking at results for KLG=2 (the threshold for radiographic osteoarthritis), mean JSW was 4.66 ±1.43 mm with a bias between operators of near zero (-0.09 mm). Patch average 95% limits of agreement (LOA) were ±0.57 mm, with limits as a percentage of the mean 12.5%, and root mean squared coefficient of variation (RMSCV) 4.3%. Better performance was noted for KLG<2, with LOA at ±0.4 mm, LOA as a percentage of the mean 7.3%, and RMSCV 2.6%. There were negligible differences in performance at the medial and lateral compartments. Viewing the presentation of results on the canonical surface, Figure 5A shows best reproducibility in the central aspect of the medial joint space, with best LOA performance of ±0.16 mm for both KLG<2 and even better at ±0.06 mm for KLG=2. Result maps for LOA as a percentage of the mean and RMSCV (Figure E2a [online])) were similar in relative distribution to LOA.

Test-retest repeatability of joint space mapping

The test-retest study included 9 individuals, 53 ± 6 years, body mass index 26 ± 4 kg/m^2, 7 women (Table 1). Results are summarized as values averaged across the whole joint and broken down by subcategories of KLG<2 and KLG=2 (Table 2). Focusing on KLG=2, mean JSW was 5.12 mm, with bias between visit 2 and visit 1 of 0.03 mm. Patch average 95% LOA were ±0.66 mm, with limits as a percentage of the mean value at 13.5%, and RMSCV 4.7%. Figure 5B shows best LOA of ±0.4 mm for KLG=2 in the central medial and lateral joint spaces, improving to near zero (±0.08 mm) for KLG<2. These results for knees with structural disease are slightly worse than reproducibility and also more heterogeneous across the joint space, most likely from sensitivity of the technique to repositioning of the knee joint in a weight-bearing position for these individuals. Result maps for LOA as a percentage of the mean and RMSCV were similar in relative distribution to LOA (Figure E2b [online])).

Feasibility of joint space mapping

The feasibility study included 66 knees from 33 participants, mean age 57 ± 7 years, body mass index 27 ± 6 kg/m^2, 23 women (Table 1). Statistical parametric mapping results for JSW dependence on KLG are presented in Figure 6 alongside mean maps for each grade, revealing a region of in the

central to posterior aspect of the medial joint space where JSW was narrower by up to 0.5 mm for each increment in KLG ($p<.05$).

Personalized joint space mapping

Figure 7 shows the comparison of baseline and 24-month JSM output in the case of a 70-year-old female (age at baseline) with body mass index 35.3 kg/m^2, demonstrating that nearly 20% of posterior lateral compartment had become narrower than baseline JSW by 24 months. The smallest detectable difference mask confirms that the ~0.1mm narrowing across nearly all of the lateral compartment is within the smallest detectable difference of the technique for individuals with KLG=2.

DISCUSSION

Weight-bearing CT is an evolving technology that can capture X-ray based three-dimensional (3-D) imaging datasets of both knees simultaneously in a fixed-flexion position akin to standard radiographic views(11,16). There has been prior investigation into 3-D joint space width (JSW) measurement derived from this type of imaging (9,11), but to the best of our knowledge this is the first work to measure JSW directly from the imaging data (rather than the distance between objects created from bone segmentation) and to analyze JSW in 3-D over a knee joint surface rather than using a unidimensional reduction of JSW across a whole compartment or subregion. This approach not only removes inaccuracies that might be introduced from operator bone segmentation but also allows spatial variation in results to be visualized and 3-D statistical analysis to be performed with statistical parametric mapping, as we demonstrate.

Both reproducibility and repeatability studies show best limits of agreement below ±0.1 mm in the central medial and lateral joint spaces, a smallest detectable difference better than ±0.2 mm previously reported for radiographic minimum JSW measurement(17,18). As has been previously demonstrated at the hip(19), this technique does show some drop off in performance towards the margins of the joint space, a result of the human operator perimeter definition step that we are now looking to automate. The excellent reproducibility between an expert and novice for both KLG<2 and KLG=2 is evidence that joint space patch segmentation can be easily learned, while it is encouraging to note repeatability results were achieved without the use of a specialized SynaFlexer positioning device. As with all measurement techniques, re-testing and reporting of reproducibility and repeatability is recommended within the setting of any future applications, particularly in the context of established structural disease where test-retest repeatability was worst. This particular

results also suggests that the support of a device such as the SynaFlexer frame would be of most value for helping individuals with structural disease to maintain standardized positioning between visits; we would recommend this in all future prospective knee weight-bearing CT studies. Statistical parametric mapping feasibility showed that 3-D JSW data can be analyzed using a surface-based approach to look at the spatial relationship between experimental variables (here Kellgren and Lawrence grade, but equally could be pain or functional measures with the appropriate accompanying data). However, this particular study was limited by small numbers and the bias of having a dependent variable of JSW linked to albeit independently assessed Kellgren and Lawrence grade. Nonetheless, once a significance threshold ($p < .05$) region of interest has been established, a single summary value from this region (such as a mean, minimum, or maximum) could be used to establish diagnostic and prognostic accuracy in disease prediction models(19).

Comparing baseline and follow-up imaging on the canonical surface is also able to highlight if there has been any deterioration beyond a critical threshold value, a concept translatable to the clinic that could enhance understanding of disease patterns and progression for both clinician and patient. A personalized approach is also able to localize where JSW is changing (narrowing or widening) beyond the smallest detectable difference for the technique, 3-D information that would not be captured by looking at a single minimum JSW value, particularly if focused on medial compartment disease only. This enables a personalized approach to imaging progression of osteoarthritis that encompasses different disease phenotypes and is not restricted to medial compartment joint space narrowing.

Further study is now needed to test the diagnostic and prognostic ability of 3-D JSW distribution against important outcome measures such as patient-reported pain, functionality, and relevant clinical events such as total knee replacement. We will now embark on this in much larger numbers of Multicenter Osteoarthritis Study participants with baseline and 2-year follow-up weight-bearing knee CT for exactly these purposes, also comparing the performance of CT against radiographic and MRI measures.

Conclusion

This study reports on three-dimensional) quantitative analysis of knee weight-bearing CT with joint space mapping, a process highly relevant to the assessment of structural joint diseases such as osteoarthritis. Reproducibility and repeatability results show a best performance of less than ±0.1 mm in the central joint spaces, at least 50% improvement on the smallest detectable difference previously reported for radiographic minimum knee joint space width measurement. Statistical

parametric mapping feasibility results show that a three-dimensional surface-based approach to analyzing joint space width can demonstrate significant relationships with structural disease. Joint space mapping is also reliably learned and performed by novice users. The lower dose of cone beam CT compared to standard clinical CT also alleviates concerns over radiation exposure when considering repeat exposures, meaning that this approach can be justified for research studies and evaluation in the clinic, with a personalized follow-up approach also possible for individuals.

REFERENCES

1. Osteoarthritis: Structural Endpoints for the Development of Drugs, Devices, and Biological Products for Treatment Guidance for Industry. 10001 New Hampshire Ave., Silver Spring, MD, USA: Center for Drug Evaluation and Research, Food and Drug Administration; 2018 p. 6. https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071577.pdf.
2. Palmer AJR, Brown CP, McNally EG, et al. Non-invasive imaging of cartilage in early osteoarthritis. Bone Jt J. 2013;95-B(6):738–746. doi: 10.1302/0301-620X.95B6.31414.
3. Eckstein F, Wirth W, Lohmander LS, Hudelmaier MI, Frobell RB. Five-year followup of knee joint cartilage thickness changes after acute rupture of the anterior cruciate ligament. Arthritis Rheumatol Hoboken NJ. 2015;67(1):152–161. doi: 10.1002/art.38881.
4. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage. 2011;19(8):990–1002. doi: 10.1016/j.joca.2011.05.004.
5. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage. 2007;15 Suppl A:A1-56. doi: 10.1016/j.joca.2006.11.009.
6. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494–502.
7. Ornetti P, Brandt K, Hellio-Le Graverand M-P, et al. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage. 2009;17(7):856–863. doi: 10.1016/j.joca.2009.01.007.
8. Bousson V, Lowitz T, Laouisset L, Engelke K, Laredo J-D. CT imaging for the investigation of subchondral bone in knee osteoarthritis. Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA. 2012;23 Suppl 8:S861-865. doi: 10.1007/s00198-012-2169-5.
9. Segal NA, Frick E, Duryea J, et al. Comparison of Tibiofemoral Joint Space Width Measurements from Standing CT and Fixed Flexion Radiography. J Orthop Res Off Publ Orthop Res Soc. 2017;35(7):1388–1395. doi: 10.1002/jor.23387.
10. Turmezei TD, Treece GM, Gee AH, et al. Quantitative 3D imaging parameters improve prediction of hip osteoarthritis outcome. Sci Rep. 2020;10(1):4127. doi: 10.1038/s41598-020-59977-2.
11. Segal NA, Bergin J, Kern A, Findlay C, Anderson DD. Test-retest reliability of tibiofemoral joint space width measurements made using a low-dose standing CT scanner. Skeletal Radiol. 2017;46(2):217–222. doi: 10.1007/s00256-016-2539-8.

12. Saltybaeva N, Jafari ME, Hupfer M, Kalender WA. Estimates of effective dose for CT scans of the lower extremities. Radiology. 2014;273(1):153–159. doi: 10.1148/radiol.14132903.
13. Bland JM, Altman DG. Agreement Between Methods of Measurement with Multiple Observations Per Individual. J Biopharm Stat. Taylor & Francis; 2007;17(4):571–582. doi: 10.1080/10543400701329422.
14. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp. 1994;2(4):189–210. doi: 10.1002/hbm.460020402.
15. Gee AH, Treece GM. Systematic misregistration and the statistical analysis of surface data. Med Image Anal. 2014;18(2):385–393. doi: 10.1016/j.media.2013.12.007.
16. Tuominen EKJ, Kankare J, Koskinen SK, Mattila KT. Weight-bearing CT imaging of the lower extremity. AJR Am J Roentgenol. 2013;200(1):146–148. doi: 10.2214/AJR.12.8481.
17. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24(1):27–67. doi: 10.1177/0962280214537344.
18. Hunter DJ. Risk stratification for knee osteoarthritis progression: a narrative review. Osteoarthritis Cartilage. 2009;17(11):1402–1407. doi: 10.1016/j.joca.2009.04.014.
19. Turmezei TD, Treece GM, Gee AH, Houlden R, Poole KES. A new quantitative 3D approach to imaging of structural joint disease. Sci Rep. 2018;8(1):9280. doi: 10.1038/s41598-018-27486-y.

TABLES

Table 1 Patient demographics

	Reproducibility	Repeatability	Feasibility
Number of individuals	20	10	33
Age (years)	58 ± 7	53 ± 6	57 ± 7
BMI	28 ± 6	26 ± 4	27 ± 6
Number of women	14	7	23

Table 2 Patch reproducibility and repeatability metrics for JSW according to KLG.

	Reproducibility KLG<2	Reproducibility KLG=2	Repeatability KLG<2	Repeatability KLG=2
Number of knees	32	6	7	7
Number of pairs	16	2	2	2
Mean ± SD* (mm)	5.54 ± 1.49	4.66 ± 1.43	4.66 ± 1.06	5.12 ± 1.20
Bias (mm)	-0.01	-0.09	0	0.03
LOA* (mm)	0.4	0.57	0.37	0.66
lateral | medial split	0.37|0.43	0.56|0.58	0.38|0.36	0.63|0.64
lower (95% CI)	-0.42 (-0.42,-0.41)	-0.66 (-0.68,-0.64)	-0.37 (-0.38,-0.36)	-0.63 (-0.65,-0.61)
upper (95% CI)	0.39 (0.38,0.39)	0.49 (0.47,0.51)	0.37 (0.36,0.38)	0.69 (0.67,0.71)
Best LOA (mm)**	0.16 (central medial)	0.06 (central medial)	0.08 (central medial) (central lateral)	0.40 (central medial) (central lateral)
LOA (as % of mean)	7.3%	12.5%	8.1%	13.5%
RMSCV (%)	2.6%	4.3%	2.7%	4.7%

*Bland-Altman one-way analysis of variance for calculating agreement between methods of measurement with multiple observations per individual (https://doi.org/10.1080/10543400701329422)

**The best regional LOA as determined from the canonical surface distributions in Figure 5.

Note. KLG = Kellgren and Lawrence grade; LOA = limits of agreement; RMSCV = root mean square coefficient of variation.

FIGURE CAPTIONS

Figure 1 (1) Shape-model assisted segmentation of the distal femur cortical outline using a threshold mask with 1.5 mm between segmented axial slices. (2) and (3) Three-dimensional distal femur surface automatically constructed within StradView from the contour set. (4) Display window width and level adjustment to create contrast between bone (white) and all other tissues (black): this does not affect the underlying imaging data values on which the measurement algorithm will run. (5) Automated projection of the average display value along the surface normal sample line back to each vertex, casting a "shadow" of the opposing joint surface to define a perimeter for the joint space. (6) Manual mark-up of this perimeter for extraction of the tibiofemoral joint space patch from the distal femur articular surfaces. (7) and (8) The reconstructed image data volume is sampled automatically along a line perpendicular to each vertex running through the joint space in three dimensions with deconvolution performed by a full width half maximum algorithm (Figure 3), defining the distance between the half-maximum at each bone surface as joint space width. (9) Independent measurements are repeated across all vertices in the patch and blurred across the surface, displaying a three-dimensional joint space width map of the distance between opposing articular bone surfaces. 2-D = two-dimensional; 3-D = 3-dimensional; A = anterior; JSW = joint space width; L = lateral; M = medial; P = posterior.

Figure 2 A full width half maximum model automatically defines the distance between bony joint surfaces as the joint space width. The two bone surface margins are set as the half of the maximum peak of the subchondral bone plate. To the best of our knowledge, this is the first CT-based joint space width measurement approach that uses this deconvolution on the imaging data rather than measuring the physical distance between bones surfaces as contoured by a human operator. Attenuation units in cone beam CT are nearly but not exactly equivalent to Hounsfield units. 1-D = one-dimensional; 3-D = three-dimensional; AU = attenuation units; FWHM = full width half maximum; JSW = joint space width.

Figure 3 Articular bone surfaces at the femur and tibia from joint space mapping output (yellow) with the halfway surface (orange), alongside joint space width displayed on the halfway patches (far right). This is shown for individuals with a radiographic Kellgren & Lawrence grade of 0 (top) and 4 (bottom), demonstrating robust performance in the extreme of disease. A = anterior; JSW = joint space width; KLG = Kellgren & Lawrence grade; L = lateral; M = medial; P = posterior.

Figure 4 Flowchart of participants involved in each of the sub-studies: reproducibility in blue, repeatability in yellow, and feasibility in green. SPM = statistical parametric mapping.

Figure 5 A, Reproducibility study results for Kellgren and Lawrence grade <2 and Kellgren and Lawrence grade =2 showing three-dimensional maps for the group mean (top row), standard deviation (second row), bias (third row), and limits of agreement (bottom row). B, As for A but from the repeatability study. The ability to present results on the three-dimensional canonical surface shows how they can vary spatially across the joint space: Table 2 provides patch average values from the whole joint space but cannot reveal where best performance is located. All units in mm. JSW = joint space width; KLG = Kellgren and Lawrence grade; SD = standard deviation.

Figure 6 Three-dimensional mean joint space width and standard deviation maps for each Kellgren and Lawrence category of <2, =2, and >2 along with the statistical parametric mapping result map show a significant region of joint space width dependence on Kellgren and Lawrence grade in the posterior aspect of the medial joint space. This demonstrated up to 0.5 mm of narrower joint space here for each grade increment. All units are in mm. JSW = joint space width; KLG = Kellgren and Lawrence grade; SD = standard deviation.

Figure 7 Comparison of three-dimensional joint space width maps and distribution histograms of an individual with baseline and 24-month follow-up weight-bearing CT. Not only can the joint space be visualized in three dimensions, but a threshold mask can be applied (red in the histogram and threshold map) to show where joint space width has progressed beyond the lowest baseline value. Histogram analysis yields box and whisker plots (median, interquartile range, and 1.5 x interquartile range). Threshold areas can be displayed in three dimensions at the distal femur to aid visualization. Finally, one can show a baseline-follow up difference map with regions of change beyond the smallest detectable distance (for Kellgren and Lawrence grade = 2 in this example) revealed with a mask. In this case, nearly all of the lateral compartment is within these limits. IQR = interquartile range; KLG = Kellgren and Lawrence grade; SDD = smallest detectable difference.

Figure E1 The canonical right-sided lateral and medial joint space surfaces (left) are registered to each individual's surfaces first using a similarity transform that allows manipulation of scale and positioning (middle), then a locally affine deformation (right). L = lateral; M = medial.

Figure E2 A, Three-dimensional reproducibility maps for limits of agreement as a percentage of the mean at each vertex (top row) and root mean square coefficient of variation (bottom row) for Kellgren and Lawrence grades <2 and =2. B, As for A but for repeatability. JSW = joint space width; KLG = Kellgren and Lawrence grade; RMSCV = root mean square coefficient of variation.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	✔ (N/A)
Title and abstract			
	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	✓
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	✓
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	✓
Objectives	3	State specific objectives, including any prespecified hypotheses	✓
Methods			
Study design	4	Present key elements of study design early in the paper	✓
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	✓
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	✓
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	✓
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	✓
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	✓
Bias	9	Describe any efforts to address potential sources of bias	✓
Study size	10	Explain how the study size was arrived at	✓
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	✓
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	✓
		(*b*) Describe any methods used to examine subgroups and interactions	✓
		(*c*) Explain how missing data were addressed	N/A
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	✓

10 E. Doty St., Suite 441, Madison, WI 53703, 630-481-1047, radiology@rsna.org

Results			✔ (N/A)
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	✓
		(b) Give reasons for non-participation at each stage	✓
		(c) Use a flow diagram and include the figure number (preferably figure 1) or page number	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	✓
		(b) Indicate number of participants with missing data for each variable of interest	N/A
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	✓
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	✓
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	*N/A*
		Cross-sectional study—Report numbers of outcome events or summary measures	*N/A*
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	✓
		(*b*) Report category boundaries when continuous variables were categorized	✓
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	✓
Discussion			
Key results	18	Summarise key results with reference to study objectives	✓
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	✓
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	✓
Generalisability	21	Discuss the generalisability (external validity) of the study results	✓
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	✓

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

*N/A stands for not applicable and may be a reasonable choice depending on the type of study performed

10 E. Doty St., Suite 441, Madison, WI 53703, 630-481-1047, radiology@rsna.org

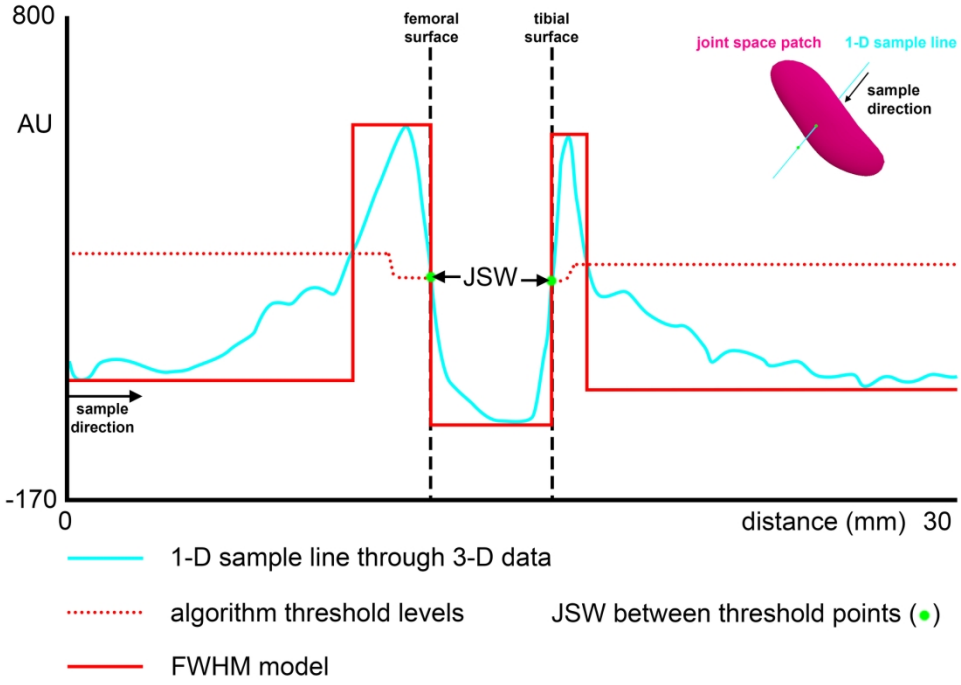

Figure 2 A full width half maximum model automatically defines the distance between bony joint surfaces as the joint space width. The two bone surface margins are set as the half of the maximum peak of the subchondral bone plate. To the best of our knowledge, this is the first CT-based joint space width measurement approach that uses this deconvolution on the imaging data rather than measuring the physical distance between bones surfaces as contoured by a human operator. Attenuation units in cone beam CT are nearly but not exactly equivalent to Hounsfield units. 1-D = one-dimensional; 3-D = three-dimensional; AU = attenuation units; FWHM = full width half maximum; JSW = joint space width.

177x132mm (300 x 300 DPI)

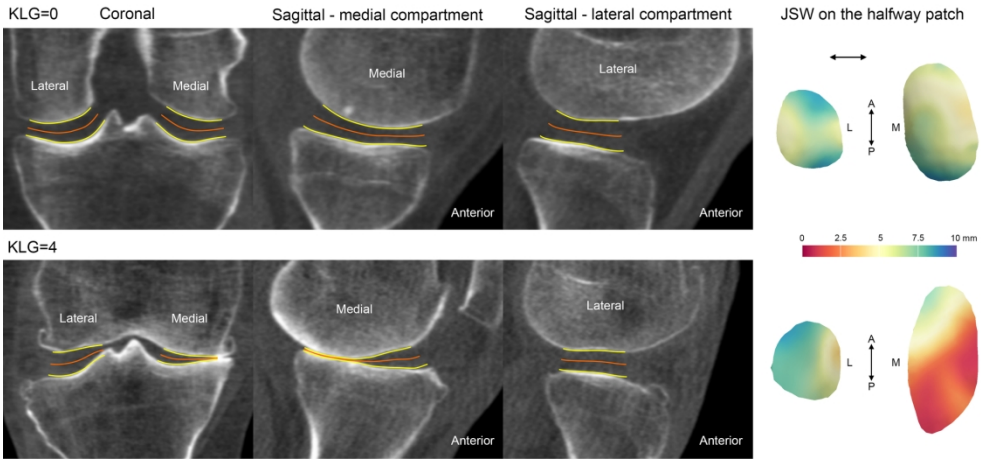

Figure 3 Articular bone surfaces at the femur and tibia from joint space mapping output (yellow) with the halfway surface (orange), alongside joint space width displayed on the halfway patches (far right). This is shown for individuals with a radiographic Kellgren & Lawrence grade of 0 (top) and 4 (bottom), demonstrating robust performance in the extreme of disease. A = anterior; JSW = joint space width; KLG = Kellgren & Lawrence grade; L = lateral; M = medial; P = posterior.

177x81mm (300 x 300 DPI)

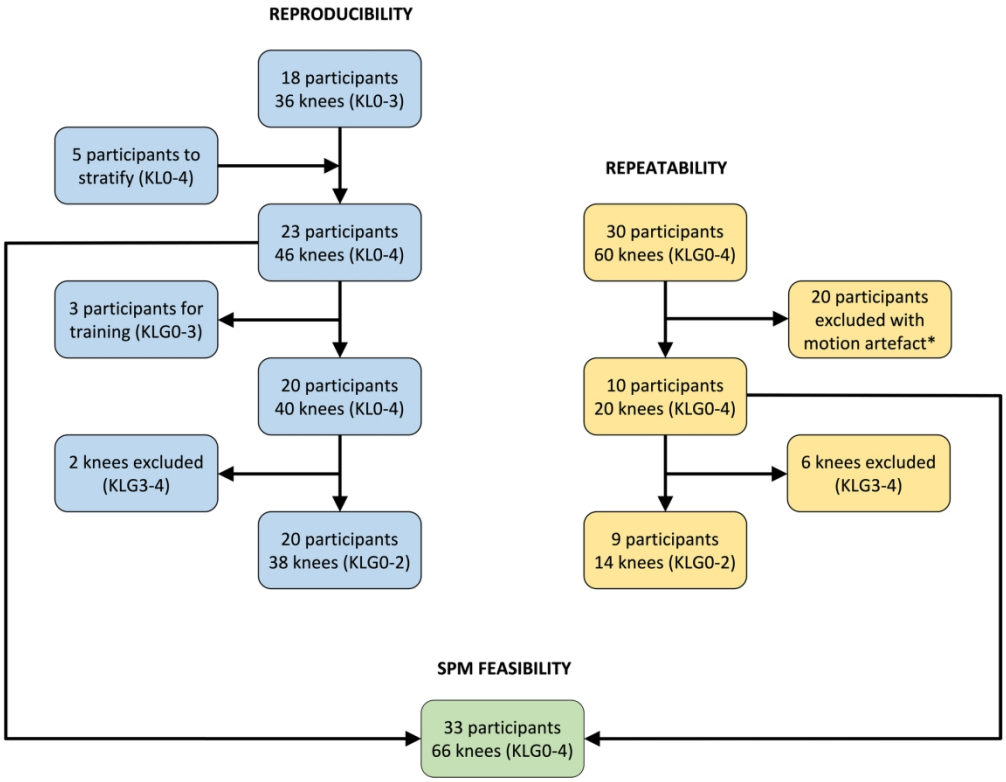

Figure 4 Flowchart of participants involved in each of the sub-studies: reproducibility in blue, repeatability in yellow, and feasibility in green. SPM = statistical parametric mapping.

177x139mm (300 x 300 DPI)

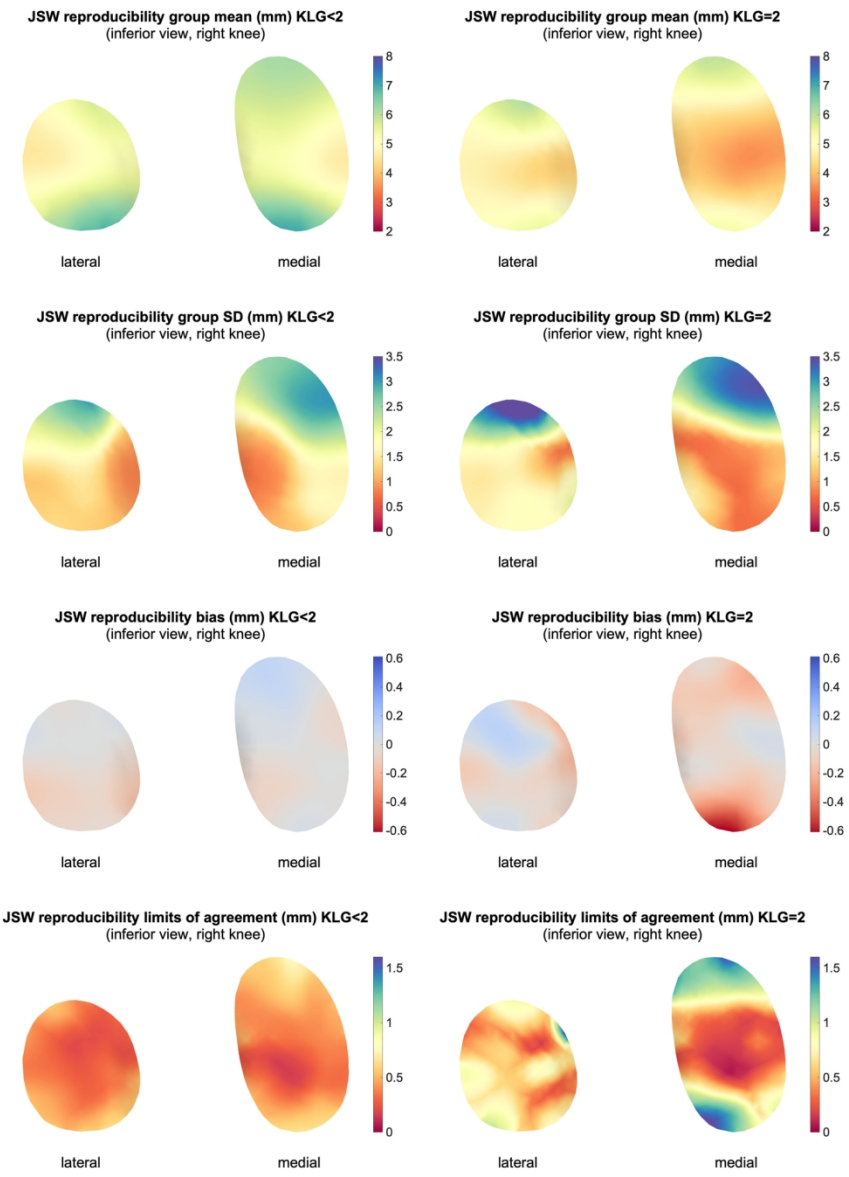

Figure 5 A, Reproducibility study results for Kellgren and Lawrence grade <2 and Kellgren and Lawrence grade =2 showing three-dimensional maps for the group mean (top row), standard deviation (second row), bias (third row), and limits of agreement (bottom row).

129x177mm (300 x 300 DPI)

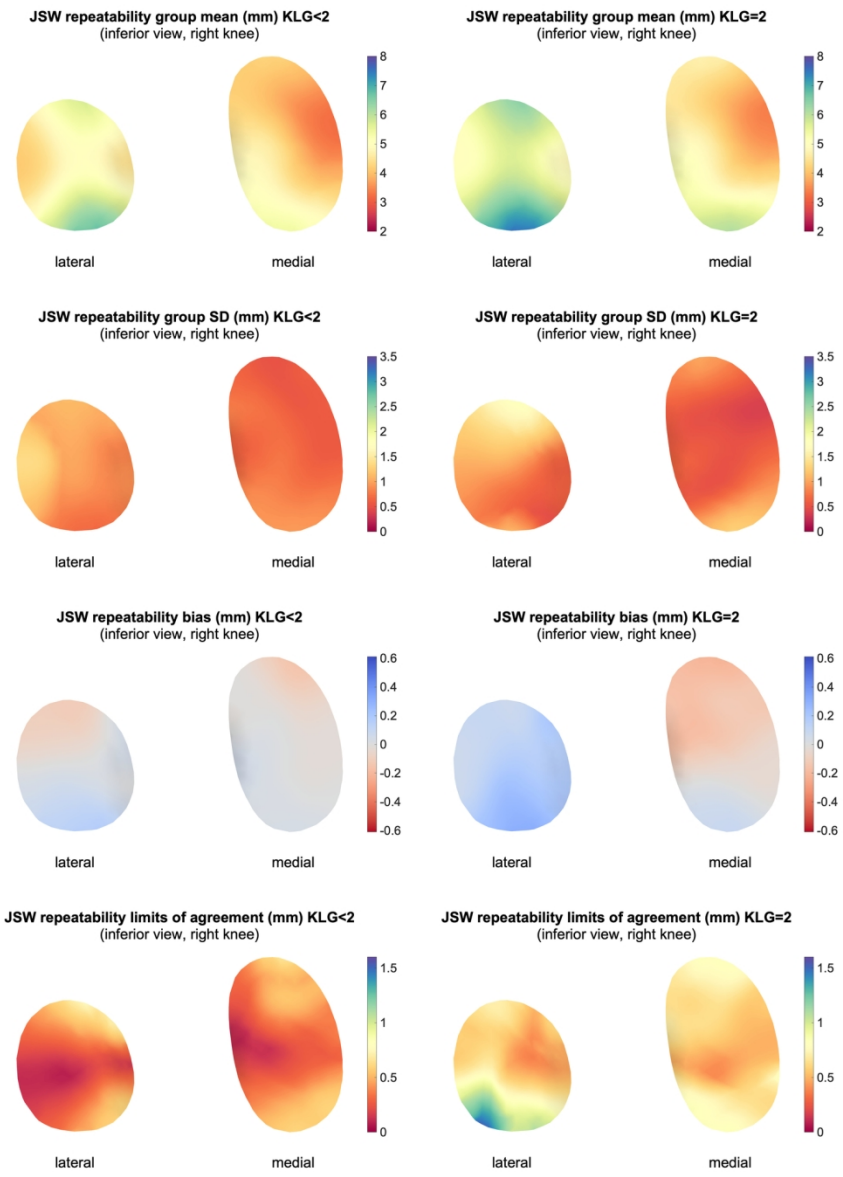

Figure 5 B, As for A but from the repeatability study. The ability to present results on the three-dimensional canonical surface shows how they can vary spatially across the joint space: Table 2 provides patch average values from the whole joint space but cannot reveal where best performance is located. All units in mm. JSW = joint space width; KLG = Kellgren and Lawrence grade; SD = standard deviation.

129x177mm (300 x 300 DPI)

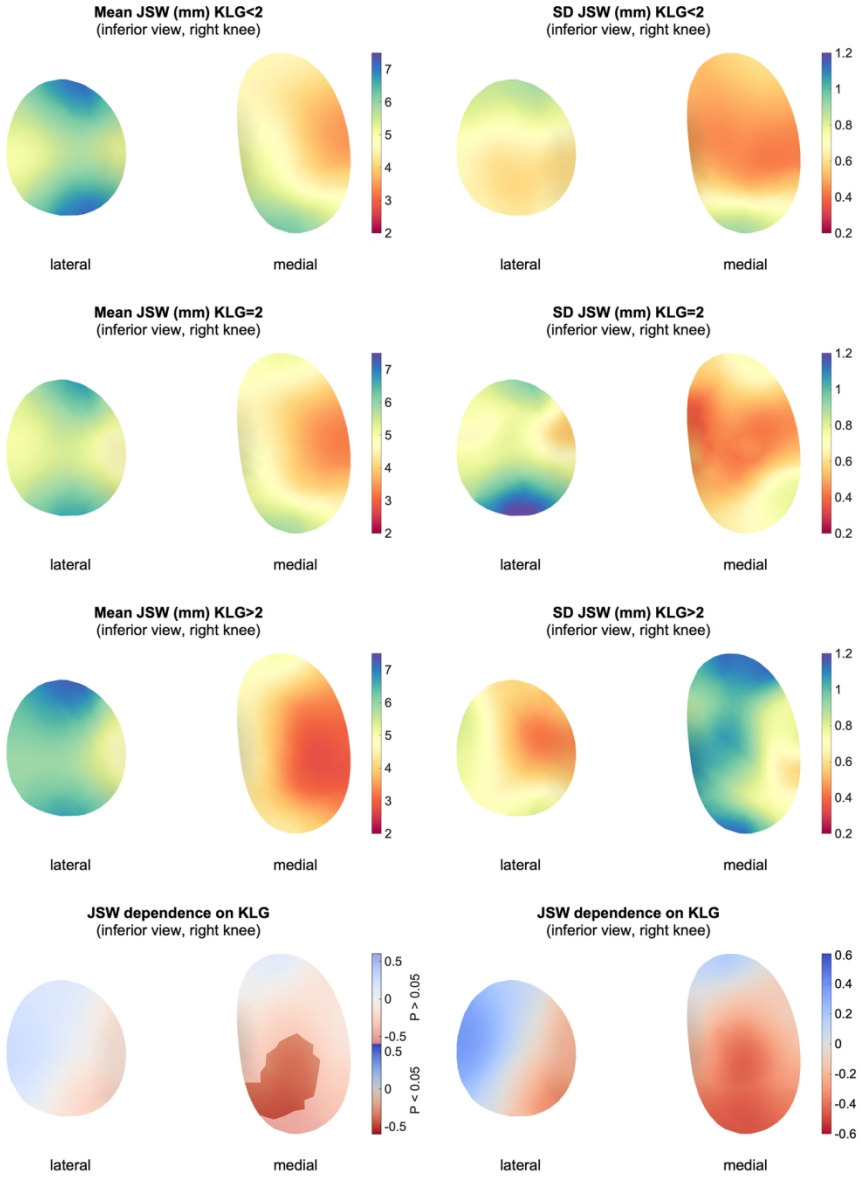

Figure 6 Three-dimensional mean joint space width and standard deviation maps for each Kellgren and Lawrence category of <2, =2, and >2 along with the statistical parametric mapping result map show a significant region of joint space width dependence on Kellgren and Lawrence grade in the posterior aspect of the medial joint space. This demonstrated up to 0.5 mm of narrower joint space here for each grade increment. All units are in mm. JSW = joint space width; KLG = Kellgren and Lawrence grade; SD = standard deviation.

133x177mm (300 x 300 DPI)

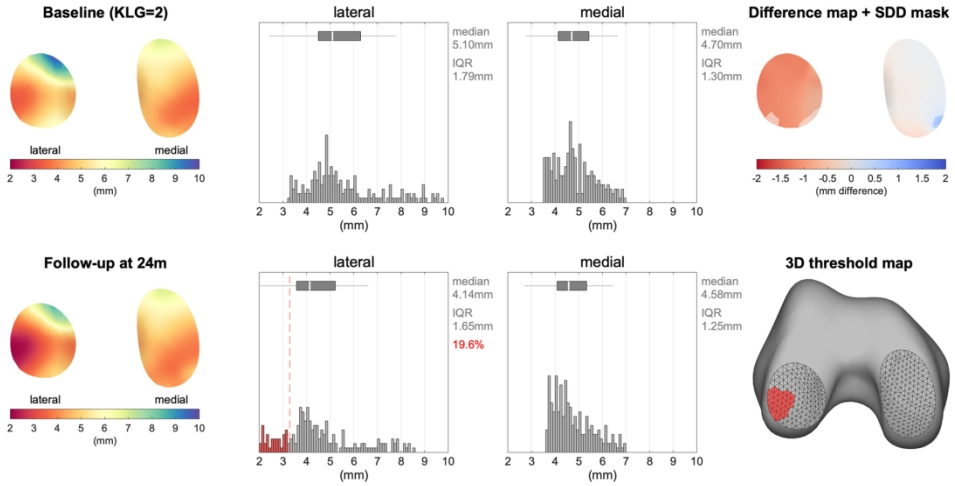

Figure 7 Comparison of three-dimensional joint space width maps and distribution histograms of an individual with baseline and 24-month follow-up weight-bearing CT. Not only can the joint space be visualized in three dimensions, but a threshold mask can be applied (red in the histogram and threshold map) to show where joint space width has progressed beyond the lowest baseline value. Histogram analysis yields box and whisker plots (median, interquartile range, and 1.5 x interquartile range). Threshold areas can be displayed in three dimensions at the distal femur to aid visualization. Finally, one can show a baseline-follow up difference map with regions of change beyond the smallest detectable distance (for Kellgren and Lawrence grade = 2 in this example) revealed with a mask. In this case, nearly all of the lateral compartment is within these limits. IQR = interquartile range; KLG = Kellgren and Lawrence grade; SDD = smallest detectable difference.

177x91mm (300 x 300 DPI)

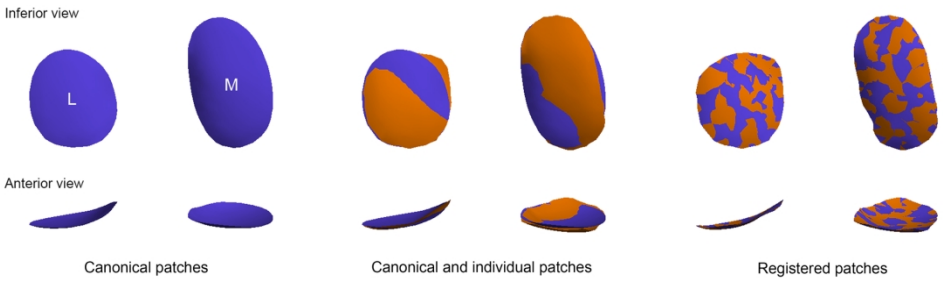

Figure E1 The canonical right-sided lateral and medial joint space surfaces (left) are registered to each individual's surfaces first using a similarity transform that allows manipulation of scale and positioning (middle), then a locally affine deformation (right). L = lateral; M = medial.

177x54mm (300 x 300 DPI)

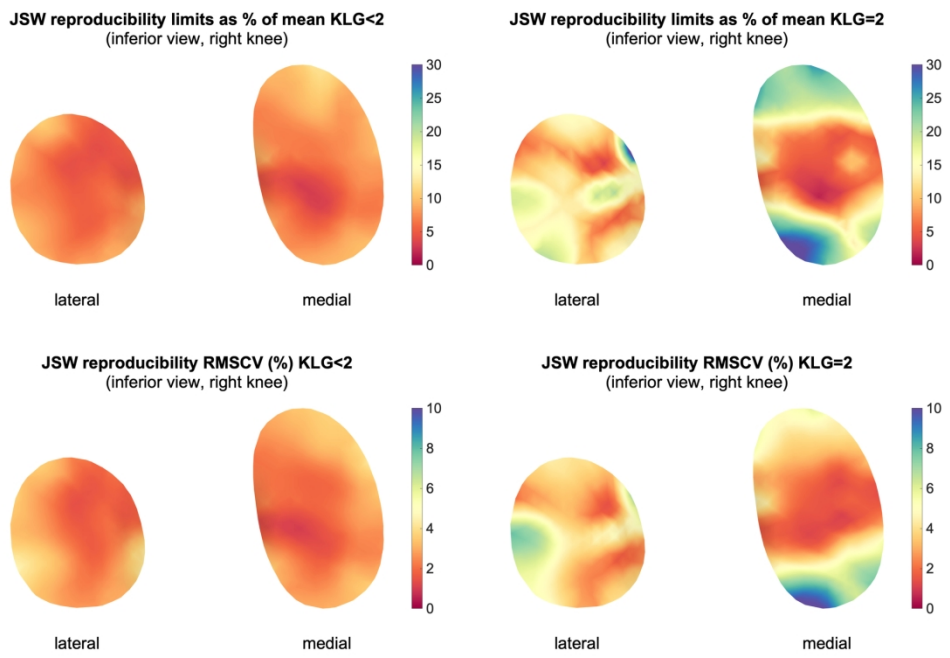

Figure E2 A, Three-dimensional reproducibility maps for limits of agreement as a percentage of the mean at each vertex (top row) and root mean square coefficient of variation (bottom row) for Kellgren and Lawrence grades <2 and =2.

177x122mm (300 x 300 DPI)

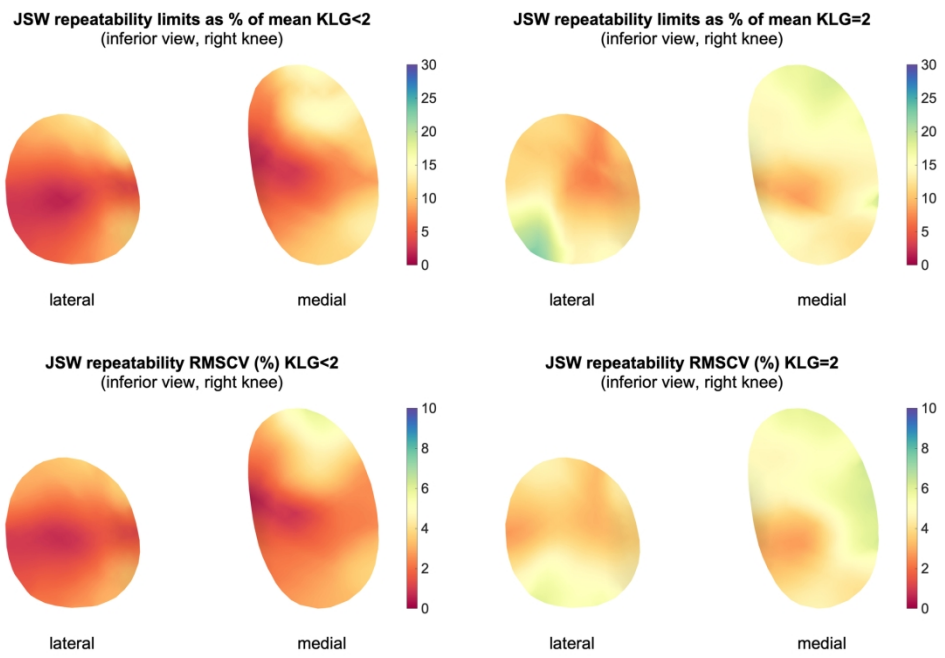

Figure E2 B, As for A but for repeatability. JSW = joint space width; KLG = Kellgren and Lawrence grade; RMSCV = root mean square coefficient of variation.

177x122mm (300 x 300 DPI)