Osteoarthritis and Cartilage 29 (2021) 1530–1539

Osteoarthritis and Cartilage

Performance of knee image digital analysis of radiographs of patients with end-stage knee osteoarthritis

M.P. Jansen [†] *, P.M.J. Welsing [†], K.L. Vincken [‡], S.C. Mastbergen [†]

[†] Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, the Netherlands
[‡] Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 24 February 2021
Accepted 24 July 2021

Keywords:
Knee osteoarthritis
Radiography
Analysis
Intra-observer
Severe

SUMMARY

Objective: Knee Image Digital Analysis (KIDA) is standardized radiographic analysis software for measuring osteoarthritis (OA) characteristics. It was validated in mild OA, but used for severe OA as well. The current goal was to evaluate the performance of KIDA in severe OA.

Design: Of 103 patients, standardized radiographs were performed before and one and 2 years after treatment for severe OA. All radiographs were evaluated on subchondral bone density, joint space width (JSW), osteophytes, eminence height, and joint angle, twice within years by the same observer. Part of the radiographs were randomly selected for reevaluation twice within 1 month and evaluation by another observer. The intraclass correlation coefficient (ICC), smallest detectable difference (SDD) and coefficient of variation (CV) were calculated; the SDD and CV were compared to those in mild OA. The relation of severity with KIDA parameters and with observer differences was calculated with linear regression.

Results: Intra-observer ICCs were higher in the 98 severe radiographs reanalyzed within 1 month (all >0.8) than the 293 reanalyzed within years (all >0.5; most >0.8) and than inter-observer ICCs (all >0.7). SDDs and CVs were smaller when reanalyzed within a month and comparable to those in mild OA. Some parameters showed bias between readings. Severity showed significant relation with osteophytes and JSW parameters, and with the observer variation in these parameters (all $P < 0.04$).

Conclusions: KIDA is a well-performing tool also for severe OA. In order to decrease variability and SDDs, images should be analyzed in a limited time frame and randomized order.

Introduction

Osteoarthritis (OA) is a degenerative joint disease characterized by structural changes like cartilage degeneration, osteophyte formation, and subchondral bone changes[1]. In knee OA, these characteristics are usually evaluated on weight-bearing anterio-posterior or postero-anterior (PA) radiographs[2]. Although the use of imaging techniques like MRI is increasing, radiography remains the primary technique for the diagnosis and monitoring of knee OA. With the exception of joint space width (JSW) as a surrogate measure of cartilage thickness, radiography-based knee OA characteristics are most often evaluated using a grading system, such as the Kellgren & Lawrence (K&L) grade and Altman score[3,4]. While these grading systems have been validated and proven useful, stepwise scoring of OA-related parameters makes results less sensitive to small changes over time. This was one of the main motivations for the development of the Knee Images Digital Analysis (KIDA) software in 2008[5]. Using KIDA, individual radiographic knee OA features of JSW, subchondral bone density, osteophytes, tibial eminence height, and knee joint angle can be measured objectively and quantitatively resulting in continuous variables. The usefulness and validity of the Knee Image Digital Analysis (KIDA) parameters was initially demonstrated for patients with relatively mild knee OA, as indicated by their average K&L grade of 1.3, and measurements were shown to distinguish these patients from healthy controls. Indeed, such distinction in mild OA is key for early detection of presence and progression of radiographic changes. Both the inter- and intra-observer variability were proven to be relatively low, and the smallest detectable difference (SDD)

* Address correspondence and reprint requests to: M.P. Jansen, Rheumatology & Clinical Immunology, University Medical Center Utrecht, Heidelberglaan 100 (G02.223), 3584CX, Utrecht, the Netherlands.Tel: 31-88-75-517-70
E-mail addresses: m.p.jansen-36@umcutrecht.nl (M.P. Jansen), P.M.J.Welsing@umcutrecht.nl (P.M.J. Welsing), K.Vincken@umcutrecht.nl (K.L. Vincken), S.Mastbergen@umcutrecht.nl (S.C. Mastbergen).

https://doi.org/10.1016/j.joca.2021.07.013

for the different parameters showed good results as well[5]. Since then, KIDA has been used in many studies, including observational cohorts with patients with relatively mild knee OA, such as the CHECK and APPROACH cohorts, but in patients with significantly more severe OA as well[6,7,16–25,8,26–30,9–15]. End-stage OA patients treated with knee joint distraction (KJD) or high tibial osteotomy (HTO) were evaluated with KIDA before and up to 9 years after treatment [24,26,31,32]. For these end-stage knee OA cases KIDA has not been evaluated. The goal of the present study was to evaluate the performance of KIDA in patients with end-stage knee OA.

Methods

Patients

Patients were included from three clinical studies. Twenty patients with end-stage knee OA, in regular care indicated for total knee arthroplasty (TKA) and relatively young (age <60 years), were included in an open prospective study and treated with KJD. In a randomized controlled trial (RCT), where KJD was compared with TKA, 20 end-stage knee OA patients indicated for TKA were treated with KJD. In a separate RCT, KJD was compared with HTO, and another 22 and 45 patients indicated for HTO were treated with KJD or HTO, respectively. For all patients, the K&L grade was determined before treatment.

All details with regard to inclusion criteria and treatment have been described in detail previously[7,21,22,33]. All trials were approved by the medical ethical review committee of the University Medical Center Utrecht (protocol numbers 04/086, 10/359/E, and 11/072) and registered in the Netherlands Trial Register (trial numbers NL419, NL2680, and NL2761). All patients gave written informed consent, which included further use of their data for additional research.

Radiography

Standardized, semi-flexed PA radiographs were performed under full weight bearing according to the Buckland–Wright protocol[34,35]. An aluminum step wedge was placed alongside the knee, against the detector and within the field of exposure, to quantify bone density and determine the pixel size corrected for possible magnification. Radiographs were taken pre-treatment (baseline) and one and 2 years post-treatment.

KIDA analysis

The KIDA analysis method has not changed since the original publication in 2008.[5]

First the aluminum step wedge is identified by the user by indicating the four corners of the wedge, after which the program automatically draws the outline of the entire wedge and the different steps (Fig. 1). From this, it calculates the pixel size and the reference mm aluminum equivalent (mm Al eq) with which subchondral bone density can be expressed. Next, the user places a framework of four lines around the joint, that touch on the medial and lateral side of the joint (two longest vertical lines), and on the distal femur and proximal tibia (horizontal lines; Fig. 1). From these last two lines, perpendiculars are calculated, four on each area (medial and lateral femur and tibia) at pre-defined calculated positions; one circle along each perpendicular can be moved by the user to place the edge of the circle at the bone-'cartilage' interface (16 smallest circles in Fig. 1). The distance between each pair of circles is calculated to measure the JSW in mm, at four locations of the medial and the lateral compartment. These four distances can be averaged to obtain a mean medial and mean lateral JSW, and all eight distances can be averaged for a mean JSW of the whole joint, all in mm. The mean intensity in each circle is calculated as well, and can be averaged to obtain the subchondral bone density at the medial and lateral tibia and femur, expressed in mm Al eq.

The height of the medial and lateral tibial eminence is determined by placing two circles on the top of the eminences; the program calculates the distance in mm from the bottom of these circles to the line at the proximal tibia (Fig. 1). Next, the user positions four circles, one at each corner of the joint, following the original bone lines (Fig. 1). The user then indicates the outer osteophyte borders; only the borders within a quadrant (blue/green lines in Fig. 1) are included. The program then calculates the osteophyte area in mm^2 for each of the four areas (yellow in Fig. 1). Using the middle eight small circles at the bone-cartilage interface, a new line is generated for both the bone edges of the femur and the tibia separately (not displayed in Fig. 1); these two lines are used to calculate the joint angle in degrees. A negative angle indicates medial joint space narrowing. Lastly, the program gives a vertical line at the narrowest point between these two lines, within the joint edges, suggesting the location of the minimum JSW. Since the bone edges are not fully straight, the user can manually adjust the lines to indicate the actual minimum JSW (this does not affect the joint angle). The program then calculates the distance between the two horizontal lines at the location of the vertical line as a measure of minimum JSW in mm.

Additional details can be found in the original publication.[5]

Data collection

Since the first KIDA publication in 2008, all KIDA evaluations have been performed by the same observer. The radiographs from

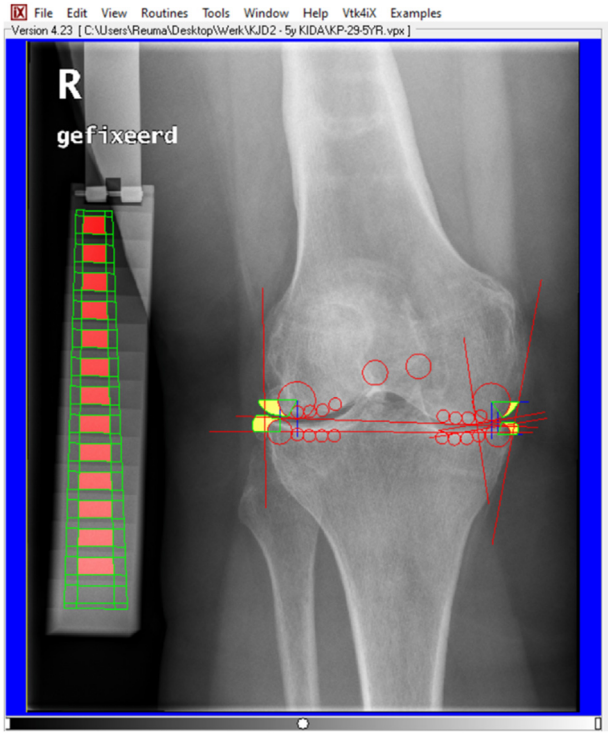

Fig. 1 Example of KIDA analysis.

Osteoarthritis and Cartilage

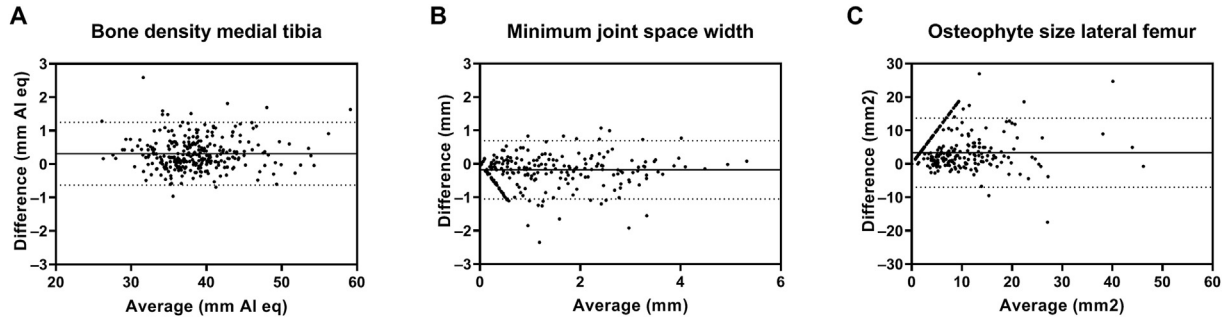

Fig. 2 Bland–Altman plots for all 293 available radiographs that were analyzed twice, for (A) the bone density of the medial tibia in mm aluminum equivalent (mm Al eq) (B) the minimum joint space width in mm and (C) the osteophyte area of the lateral femur in mm^2.

the three previously performed studies (described under *patients*) evaluated in the current analysis were all analyzed for the first time between 2013 and 2015. More recently (2017–2018) all radiographs at baseline, one and 2 years for these three studies, were reanalyzed by the same reader. As such, almost all radiographs had duplicate readings, which could be used for determining the intra-observer variability and as such for an evaluation KIDA performance in these patients with end-stage OA.

Since there were multiple years between the first and second analysis, which might influence results, 100 of the radiographs were randomly selected to be evaluated again twice with maximum 1 month in between. The selection was made randomly to ensure that the subset was generalizable to the full set of radiographs. These 200 images (100 radiographs analyzed twice) were randomly ordered and divided in four batches of 50; every week one batch was analyzed by the same observer (MM) blinded to patient characteristics. This data set was additionally used to make a comparison with the dataset from the original KIDA publication, which consisted of mild OA patients with duplicate readings with limited timespan in between both readings[5]. Moreover, the relevance of the in-between reading time, months versus years, could be evaluated.

Inter-observer variability analyses was performed as well. The sample size for this analyses was calculated using the lowest intraclass correlation coefficient (ICC) of the intra-observer experiments where images were analyzed twice within a month, a precision of 0.1, a confidence level of 95%, and 2 raters. The resulting number of images was taken from the radiographs that were analyzed twice within a month by the first observer, and analyzed by a second observer as well.

Parameter	Mean ± SD	Mean Δ	SD Δ	95%CI Δ	SDD	ICC	95%CI ICC
Bone density (all in mm Al eq)							
Femur mean lateral	33.0 ± 4.6	−0.05	0.31	−0.09–−0.02	0.61	0.998	0.997–0.998
Femur mean medial	37.3 ± 5.1	0.17	0.33	0.13–0.21	0.65	0.997	0.995–0.998
Tibia mean lateral	33.7 ± 5.2	−0.59	1.41	−0.75–−0.43	2.76	0.958	0.931–0.972
Tibia mean medial	38.6 ± 5.0	0.32	0.50	0.26–0.38	0.98	0.993	0.981–0.997
JSW (all in mm)							
Mean	5.2 ± 1.1	0.37	0.59	0.30–0.44	1.16	0.821	0.616–0.901
Mean lateral	7.7 ± 1.9	0.66	1.03	0.54–0.77	2.02	0.816	0.604–0.899
Mean medial	2.7 ± 1.7	0.09	0.39	0.04–0.13	0.76	0.973	0.965–0.979
Minimum	1.0 ± 1.1	−0.18	0.45	−0.23–−0.13	0.88	0.915	0.870–0.942
Osteophytes (all in mm^2)							
Femur lateral	7.6 ± 7.5	3.32	5.26	2.72–3.93	10.31	0.716	0.459–0.833
Femur medial	7.1 ± 8.0	5.44	6.84	4.66–6.23	13.41	0.579	0.184–0.764
Tibia lateral	10.0 ± 10.0	2.49	4.77	1.94–3.03	9.35	0.867	0.755–0.919
Tibia medial	7.6 ± 5.9	2.46	6.24	1.74–3.18	12.23	0.532	0.396–0.637
Other (mm, mm, degrees)							
Eminence lateral	12.5 ± 2.3	0.59	1.13	0.46–0.72	2.21	0.864	0.752–0.917
Eminence medial	13.3 ± 1.8	0.30	0.94	0.19–0.41	1.84	0.860	0.809–0.895
Joint angle	−6.1 ± 3.4	−0.66	1.19	−0.80–0.52	2.33	0.924	0.842–0.957

ICC = intraclass correlation coefficient; Mean ± standard deviation (SD) = mean and SD of all radiographs; mean Δ = mean difference between the two observations of all radiographs; SD = standard deviation of mean differences between the two observations; 95%CI = 95% confidence interval of mean differences between the two observations; SDD = smallest detectable difference (1.96*SD of mean differences between the two observations); mm Al eq = minimum aluminum equivalent.

Table I Intra-observer results for all severe radiographs reanalyzed within a large period of time

Parameter	Mean ± SD	Mean Δ	SD Δ	95%CI Δ	SDD	ICC	95%CI ICC
Bone density (all in mm Al eq)							
Femur mean lateral	32.6 ± 4.4	0.02	0.21	−0.02–0.06	0.41	0.999	0.998–0.999
Femur mean medial	36.8 ± 5.1	0.02	0.23	−0.02–0.07	0.45	0.999	0.999–0.999
Tibia mean lateral	32.9 ± 5.2	−0.16	1.20	−0.40–0.08	2.35	0.973	0.960–0.982
Tibia mean medial	38.2 ± 5.2	0.10	0.35	0.03–0.17	0.69	0.998	0.996–0.998
JSW (all in mm)							
Mean	5.3 ± 1.1	0.07	0.39	−0.01–0.15	0.76	0.935	0.904–0.956
Mean lateral	7.7 ± 1.9	0.17	0.73	0.03–0.32	1.43	0.923	0.885–0.948
Mean medial	2.9 ± 1.8	−0.03	0.24	−0.08–0.02	0.47	0.992	0.987–0.994
Minimum	1.0 ± 1.2	−0.04	0.31	−0.10–0.02	0.61	0.965	0.948–0.976
Osteophytes (all in mm^2)							
Femur lateral	8.6 ± 7.3	0.80	2.73	0.25–1.34	5.35	0.928	0.890–0.953
Femur medial	11.1 ± 9.1	2.04	5.69	0.90–3.18	11.15	0.806	0.702–0.873
Tibia lateral	11.8 ± 12.2	0.77	5.90	−0.41–1.95	11.56	0.889	0.838–0.924
Tibia medial	8.5 ± 7.4	0.01	5.58	−1.11–1.13	10.94	0.751	0.649–0.826
Other (mm, mm, degrees)							
Eminence lateral	12.7 ± 2.5	0.17	1.14	−0.06–0.39	2.23	0.901	0.856–0.933
Eminence medial	13.3 ± 1.8	0.06	0.89	−0.12–0.24	1.74	0.891	0.842–0.926
Joint angle	−6.0 ± 3.6	−0.28	0.96	−0.47–−0.08	1.88	0.962	0.941–0.975

ICC = intraclass correlation coefficient; Mean ± standard deviation (SD) = mean and SD of all radiographs; mean Δ = mean difference between the two observations of all radiographs; SD = standard deviation of mean differences between the two observations; 95%CI = 95% confidence interval of mean differences between the two observations; SDD = smallest detectable difference (1.96*SD of mean differences between the two observations); mm Al eq = minimum aluminum equivalent.

Table II Intra-observer results for severe radiographs reanalyzed within 1 month

Osteoarthritis and Cartilage

Statistical analysis

The intra-observer variability was calculated for the two groups of end-stage radiographs separately: the total group with radiographs analyzed with a larger and varying time period (years) between two observations, and the 100 radiographs analyzed within 1 month.

The intra-observer variation was, for all KIDA parameters separately, displayed with Bland–Altman plots in which the difference between the first and second result was plotted against the mean of the two observations[36]. The mean and standard deviation (SD) of all measurements were calculated, as were the mean, SD and 95% confidence interval (95%CI) of the differences between the duplicate readings; the SDD was defined as 1.96 times the SD of the differences. The ICC was calculated for single measures using a two-way random model with absolute agreement. ICCs were interpreted according to the definitions of Koo and Li: an ICC <0.50 was considered poor, 0.50 < ICC >0.75 was moderate, 0.75 < ICC >0.90 was good, and ICC >0.90 was excellent.[37]

The mean, SD (of the difference), and SDD were compared between the three groups of radiographs: total group with severe OA radiographs analyzed with a larger period between two observations, the 100 severe OA radiographs analyzed within 1 month, and the results from the mild OA patients from the original publication. Since the SD and with that the SDD may depend on the mean absolute values which may influence the comparison of these absolute values between mild and severe OA patients, the coefficient of variation (CV), a measure expressing variability relatively to the average value of the measurements, was calculated as well, by dividing the SD of the differences between observations by the mean value of both observations and multiplying that by 100 (%).

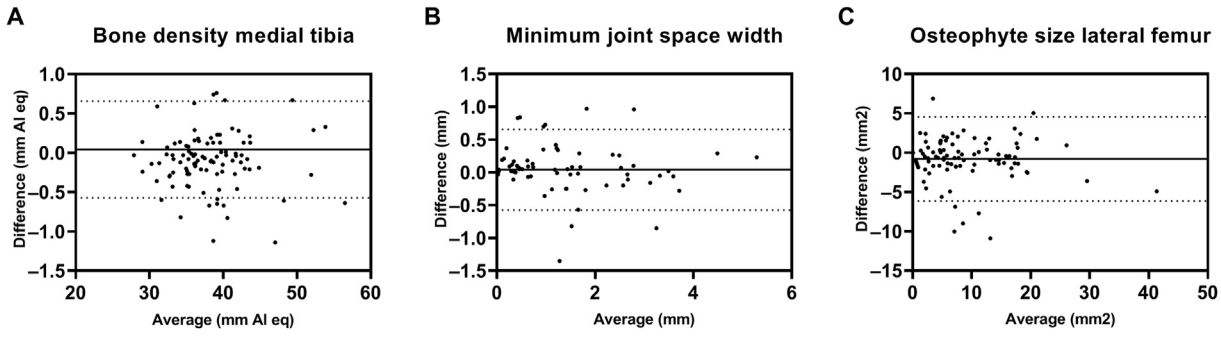

Fig. 3 Bland–Altman plots the 98 radiographs that were analyzed twice within a month, for (A) the bone density of the medial tibia in mm aluminum equivalent (mm Al eq) (B) the minimum joint space width in mm and (C) the osteophyte area of the lateral femur in mm^2.

Osteoarthritis and Cartilage

To compare KIDA parameters with the most frequently used grading system for OA, individual KIDA parameters were compared to the overall K&L grade. This was done using separate linear regression models, using only one (the most recent) analysis result for each of the radiographs analyzed within years.

As an additional explorative analysis, the influence of the mean of the measurements (of the two observations) and K&L grade (separately, as measure for severity) on the absolute intra-observer difference between two measurements was analyzed for all parameters. For this, linear regression was used on the data of severe OA patients re-analyzed within 1 month (to ensure that results will not be biased by a long period of time between analyses); a p-value <0.05 was considered statistically significant.

For the inter-observer variability, the ICC and mean difference with 95%CI were calculated for the results of the second observer compared with both the first and second analysis of the first observer separately. The influence of the mean values and K&L grade on the absolute inter-observer differences was analyzed as well.

Results

Patients

In total, 293 radiographs with double KIDA readings were available, taken at baseline ($n = 103$), 1-year follow-up ($n = 98$), and 2 year follow-up ($n = 92$). The radiographs were taken of 103 different patients, of whom 61 were treated with KJD and 42 with HTO. The mean K&L grade of the images was 2.7. The median time difference between the first and second analysis was 50 months (interquartile range 39–52 months).

Of the 100 radiographs that were reanalyzed within 1 month, for two radiographs the process of randomization was not correct and, as a result, they were not included for analysis twice. As such, these were excluded, and double analysis results within 1 month were available for a total of 98 radiographs. These were images of patients treated with KJD ($n = 56$) and HTO ($n = 42$) and taken at baseline ($n = 37$), 1 year ($n = 29$) and 2 years ($n = 32$). The average K&L grade was 2.6.

Intra-observer results for all end-stage radiographs reanalyzed within a large period of time.

The Bland–Altman plots for three relevant example parameters evaluated in all 293 radiographs are shown in Fig. 2. Plots for all other parameters of these patients can be found in Supplementary Figs. S1–5.

All bone densities [Fig. 2(A) and S1], eminence height (Fig. S4), and joint angle (Fig. S5) plots did not show large systematic differences between the two readings. However, the minimum JSW [Fig. 2(B) and S2] and osteophyte plots [Fig. 2(C) and S3] showed a floor effect, where measurements resulted in 0 in one analysis, but not in the other (as indicated by the straight line of dots starting from around 0).

The ICC, mean and SD of all measurements, mean of the differences between two analysis moments, SD and 95%CI of the difference, and the SDD are shown in Table I for all parameters. The

Parameter	Mean			SDD			CV (%)		
	Severe OA n = 293	Severe OA 1 month n = 98	Mild OA n = 55	Severe OA n = 293	Severe OA 1 month n = 98	Mild OA n = 55	Severe OA n = 293	Severe OA 1 month n = 98	Mild OA n = 55
Bone density (all in mm Al eq)									
Femur mean lateral	33.0	32.6	28.6	0.61	0.41	1.08	0.9	0.6	1.9
Femur mean medial	37.3	36.8	29.8	0.65	0.45	0.84	0.9	0.6	1.4
Tibia mean lateral	33.7	32.9	29.6	2.76	2.35	1.06	4.2	3.6	1.8
Tibia mean medial	38.6	38.2	31.3	0.98	0.69	0.84	1.3	0.9	1.4
JSW (all in mm)									
Mean	5.2	5.3	5.1	1.16	0.76	0.86	11.3	7.4	8.6
Mean lateral	7.7	7.7	6.1	2.02	1.43	1.53	13.4	9.5	12.8
Mean medial	2.7	2.9	4.2	0.76	0.47	0.67	14.4	8.3	8.1
Minimum	1.0	1.0	2.8	0.88	0.61	0.49	45.0	31.0	8.9
Osteophytes (all in mm^2)									
Femur lateral	7.6	8.6	5.4	10.31	5.35	6.78	69.2	31.7	64.1
Femur medial	7.1	11.1	3.7	13.41	11.15	3.21	96.3	51.3	44.3
Tibia lateral	10.0	11.8	6.4	9.35	11.56	8.06	47.7	50.0	64.2
Tibia medial	7.6	8.5	9.9	12.23	10.94	4.63	82.1	65.6	23.8
Other (mm, mm, degrees)									
Eminence lateral	12.5	12.7	10.0	2.21	2.23	2.47	9.0	9.0	12.6
Eminence medial	13.3	13.3	11.6	1.84	1.74	1.92	7.1	6.7	8.4
Joint angle*	6.4	6.3	3.0	2.35	1.86	2.02	18.8	15.1	34.3

SDD = smallest detectable difference (1.96*standard deviation of mean differences between the two observations); CV = coefficient of variation (standard deviation of the differences between observations divided by the mean value of both observations and multiplied that by 100); mm Al eq = minimum aluminum equivalent.

* The joint angle here was defined as the absolute value (negative angles as a result of medial JSW narrowing were taken as a positive value), as this was done in the original publication.

Table III Intra-observer results for the three groups

ICCs in most cases were good-excellent and the differences (Δ), SD and SDD were small compared to the overall means. However, the osteophyte ICCs were moderate (except for the lateral tibia with a good ICC) and the differences and SDDs were relatively high compared to the mean values. Furthermore, all parameters showed a systematic difference (bias) between readings, as indicated by the 95%CI of the difference. The direction of this bias differed and, except for osteophytes, was small relative to the absolute value.

Intra-observer results for end-stage radiographs reanalyzed within 1 month.

The Bland–Altman plots for the same set of three parameters as shown in Fig. 2, but evaluated in the 98 radiographs that were reanalyzed within a month, are shown in Fig. 3. All other plots for these 98 radiographs can be found in Supplementary Figs. S6–10. For these analyses, none of the plots showed a meaningful floor effect between the two readings.

The analysis parameters for these radiographs are shown in Table II. The ICCs were excellent for most parameters, for four parameters the ICC was good, three of them being osteophyte parameters. Again, in most cases, differences (Δ), SD and SDD were small compared to the overall means for all parameters, although not for the osteophytes. Clearly fewer parameters showed significant bias in these analyses compared to the radiographs analyzed twice in a larger period of time, as indicated by the 95% CI of the differences. Similar to the observed bias for comparisons over the longer time period, the tibia medial bone density, femoral osteophytes, and mean lateral JSW showed significant positive bias (i.e., higher scores in the second measurement), while the joint angle showed negative bias.

Comparison of intra-observer results in all three groups

The mean of the parameters, the SD of the differences, the SDD and CV are shown for the three groups (all 293 radiographs with severe OA reanalyzed within a large period of time, the 98 radiographs with severe OA reanalyzed within 1 month, and 55 radiographs with mild OA from the original publication analyzed within 1 month) in Table III. Besides increasing the ICC (comparing Tables I and II), reanalyzing the severe OA radiographs within 1 month seemed to decrease the SDD and CV for almost all parameters. Furthermore, the SDD and CV for severe OA patients analyzed within 1 month were comparable to and often even better than those for mild OA patients for most parameters. Compared to mild OA, the SDD for severe OA was especially high for osteophyte parameters, although the CV, which corrects the SD for the mean overall values, was more comparable. For the tibia lateral bone density, the difference remained high in SDD and CV.

Obviously, but importantly, all variables differed between mild and the severe OA in the expected direction, severe patients having a higher bone density, a smaller JSW, larger osteophytes, and higher eminetia.

Comparison with Kellgren–Lawrence grade

The relation between all individual KIDA parameters and K&L-grade are shown in Table IV. A smaller JSW and especially higher osteophytes were significantly associated with a higher K&L-grade, as would be expected. A higher bone density in the medial femur showed a significant positive relation with K&L-grade as well.

Influence of severity on intra-observer difference

The influence of the mean values and K&L grade, both as a measure of severity, on the differences between measurements of severe radiographs reanalyzed within 1 month are shown in Table V. Both medial osteophyte parameters and the lateral tibia osteophytes showed a statistically significant influence of the mean values and of the K&L grade (all $P < 0.02$); more severe OA (higher values) corresponded with a larger difference between measurements. Additionally, the tibia medial bone density and minimum JSW showed a significant positive influence of their mean values (both $P < 0.03$), but not K&L grade (both $P > 0.32$).

Inter-observer results

The predetermined sample size was 75. The inter-observer results are summarized in Supplementary Table S1, and show only slightly lower ICCs compared to the intra-observer results in the overlapping sample of the images reanalyzed within 1 month (Table II). However, most parameters show a systematic difference between the results of the second observer and both the first and second analysis of the first observer, and somewhat larger absolute mean differences between the analyses compared to the intra-observer differences. The SDs of the differences were also larger than those of the intra-observer analyses, which was the case for the mild OA patients as described in the original KIDA article as well. Especially minimum JSW and osteophyte size showed a dependency of the mean absolute difference between readings on the mean value and K&L grade (Supplementary Table S2), as was the case for the intra-observer results. However, for the osteophytes, more severe OA seemed to result in less inter-observer variation.

Parameter	B	β	*p*-value
Bone density (all in mm Al eq)			
Femur mean lateral	0.164	0.034	0.564
Femur mean medial	1.218	0.227	<0.001
Tibia mean lateral	0.314	0.056	0.324
Tibia mean medial	0.608	0.113	0.053
JSW (all in mm)			
Mean	−0.173	−0.142	0.015
Mean lateral	−0.049	−0.023	0.697
Mean medial	−0.298	−0.163	0.005
Minimum	−0.299	−0.247	<0.001
Osteophytes (all in mm^2)			
Femur lateral	2.995	0.340	<0.001
Femur medial	4.917	0.488	<0.001
Tibia lateral	4.677	0.402	<0.001
Tibia medial	2.325	0.325	<0.001
Other (mm, mm, degrees)			
Eminence lateral	0.097	0.038	0.520
Eminence medial	−0.016	−0.008	0.891
Joint angle	−0.358	−0.096	0.100

B = unstandardized coefficient; β = standardized coefficient; mm Al eq = minimum aluminum equivalent.
Separate linear regression models were used for all different parameters.

Table IV Relation between KIDA parameters and Kellgren–Lawrence grade

Osteoarthritis and Cartilage

	Mean value			Kellgren–Lawrence grade		
Parameter	B	β	*P*-value	B	β	*P*-value
Bone density (all in mm Al eq)						
Femur mean lateral	0.002	0.078	0.443	0.024	0.176	0.083
Femur mean medial	0.006	0.186	0.066	−0.013	−0.081	0.427
Tibia mean lateral	0.032	0.159	0.118	−0.060	−0.057	0.575
Tibia mean medial	0.011	0.223	**0.021**	0.025	0.100	0.326
JSW (all in mm)						
Mean	−0.028	−0.098	0.336	−0.025	−0.081	0.428
Mean lateral	0.011	0.032	0.755	−0.066	−0.104	0.309
Mean medial	−0.006	−0.067	0.515	0.016	0.093	0.365
Minimum	0.070	0.314	**0.002**	0.008	0.029	0.778
Osteophytes (all in mm^2)						
Femur lateral	0.050	0.173	0.089	0.156	0.072	0.479
Femur medial	0.130	0.244	**0.016**	1.340	0.271	**0.007**
Tibia lateral	0.257	0.577	**<0.001**	1.783	0.325	**0.001**
Tibia medial	0.393	0.606	**<0.001**	1.375	0.285	**0.004**
Other (mm, mm, degrees)						
Eminence lateral	0.043	0.114	0.264	0.134	0.141	0.167
Eminence medial	0.021	0.059	0.564	−0.040	−0.059	0.565
Joint angle	−0.011	−0.048	0.639	−0.069	−0.084	0.413

B = unstandardized coefficient; β = standardized coefficient; mm Al eq = minimum aluminum equivalent.
Separate linear regression models were used for the mean value and the Kellgren–Lawrence grade, and for all different parameters. Bold p-values indicate statistical significance ($P<0.05$).

Table V Influence of mean values and Kellgren–Lawrence grade on the intra-observer differences between measurements for severe radiographs analyzed within 1 month

Osteoarthritis and Cartilage

Discussion

Based on the presented results, KIDA is a useful tool for radiographic analysis of OA characteristics even in patients with severe OA. Notably (re)analyzing images in a short time period increases reproducibility (decreases SDDs and CVsand decreases systematic bias between measurements). This emphasizes the importance of performing the analyses required for a specific research question within a limited time period and randomized for time/visit sequence. Ideally, the images should be analyzed by one observer, as the analysis by multiple observers results in more significant bias between measurements as well.

The fact that some parameters showed significant differences between readings, even for the images reanalyzed within 1 month, can only be speculated on. For these parameters, the direction of this bias was the same for the images that were analyzed over years and over months, which implies that the bias is expectedly systematic and not coincidental (i.e., not because of subtly different conditions that may unconsciously affect measurement) and that it is not likely the result of recalling the first reading. For most consistent biases, the direction was positive. As such, changes over time for bone density, femoral osteophytes, and lateral JSW might be overstated when images are analyzed in chronological order of acquisition over time (visits). However, the bias is small compared to the mean values and treatment effect observed thus far[24,26,31]. Moreover, this bias becomes irrelevant when comparing differences in changes over time between groups, e.g., treatment arms. Still, when analyzing changes over time, it is strongly recommended to randomize the order in which radiographs are analyzed, so this bias will not be of relevance.

Although speculative, the systematic bias for bone density and osteophyte area may be caused by a gradual learning curve of the observer in identifying the outer and inner boundaries of the osteophytes and the edges of the bone-cartilage interface (black-to-white interface on the radiographs). Moving the small circles that determine JSW and bone density somewhat may not affect JSW significantly, but if the circle is placed slightly outside the actual (white) bone area, a small number of pixels could be dark-gray to black (background) and impact the average gray value. It would be interesting to repeat the reanalysis within both a short and long time with one or more different observers in the future, to see whether similar intra-observer results are found and if the cause of the systemic bias can be determined. However, finding a definitive cause may be difficult.

It is remarkable that for many parameters, the SDD was lower (better) for severe OA patients analyzed within a month in this study than for mild OA patients from the original publication. However, the differences are not very large. Again the explanation may be found in a learning curve by analyzing KIDA images over the past 12 years. In this case the experience is in favor of the technique (reproducibility), instead of the time-dependent bias.

The more important conclusion is that for most parameters, intra-observer variation is similar in severe OA patients compared to mild OA patients. Medial osteophyte areas seem to be the exception, and have a much bigger (worse) SDD for severe OA patients. For both medial osteophytes and lateral tibial osteophytes it was shown that the intra-observer variation depended on the osteophyte area, as bigger osteophytes, associated with more severe OA, and a higher K&L grade results in a larger variation between measurements. This explains why, even if SDDs are not comparable between patients of different severities, the CVs are (as they are corrected for the mean osteophyte area). Surprisingly, an opposite effect was seen for inter-observer results, as larger osteophytes resulted in smaller differences between observers. Osteophytes did not only show a relatively high dependence on mean values, but also on whether the reanalysis was performed

within a long or short period. All four osteophyte locations showed a clear floor effect in the complete dataset of 293 radiographs (Fig. S3), which disappeared for the 98 radiographs reanalyzed within 1 month (Fig. S8). This may also be explained by a learning curve, as these osteophytes were not recognized as osteophytes in the first reading (value 0) but were recognized as osteophytes in the second reading. Furthermore, while ICC improved for all parameters when reanalyzed within a month (compare Table II with Table I), this effect was the most notable for the osteophyte measurements. Apparently the osteophytes are the parameters most sensitive to intra-observer variability. This may be explained by the fact that the values depend on a calculated area within a manually delineated boundary, a subjective action sensitive to a learning curve.

While the minimum JSW SDD is comparable between mild and severe OA, the CV shows a large difference, because the severe OA patients show a smaller mean value for minimum JSW. Surprisingly, for minimum JSW, a higher absolute difference between intra- and inter-observer measurements was significantly associated with a higher mean values (and thus less severe OA), although this comparison could have been complicated by the extremely small values, as a result of a truncation effect (one-sided limitation at 0) and limitations with respect to pixel size. Nevertheless, also in these cases, performing the analyses in a short time frame greatly decreases this variability.

The SDDs calculated in this research indicate the smallest change that can be interpreted as a real change, as opposed to a measurement error, with $P < 0.05$. It is important to note that the SDDs described in this research are relevant on an individual level. On a group level, e.g., when evaluating groups of patients before and after treatment, the group SDD should be calculated by dividing the SDDs calculated here by the square root of the number of observations in the group.[38,39]

Apart from differences between measurements, it was shown that also in more severe OA, osteophytes and JSW parameters were significantly associated with K&L grade. As such, as for mild OA, also for severe OA KIDA is a valid method to evaluate radiographic characteristics of OA.

In conclusion, while the variability of some parameters may depend on severity, and without precautions bias may develop, KIDA has been shown to be a useful tool also in patients with severe OA. Its use, like most image analyses techniques, needs to be performed with caution. In order to decrease variability and be able to detect smaller differences, images should be analyzed in a limited time frame and randomized order.

Contributions

All authors have made substantial contributions to all three of sections (1), (2) and (3) below:

(1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data
(2) drafting the article or revising it critically for important intellectual content
(3) final approval of the version to be submitted

Specifically:
Planning and design of study: MJ, PW, SM.
Data collection, analysis and interpretation: MJ, PW, KV, SM.
Drafting the article: MJ.
Critically revising the article: PW, KV, SM.

Mylène Jansen (m.p.jansen-36@umcutrecht.nl) and Simon Mastbergen (s.mastbergen@umcutrecht.nl) take responsibility for the integrity of the work as a whole, from inception to finished article.

Conflict of interest

The authors declare no conflict of interest.

Ethics approval of research on humans

All trials were approved by the medical ethical review committee of the University Medical Center Utrecht (protocol numbers 04/086, 10/359/E, and 11/072) and registered in the Netherlands Trial Register (trial numbers NL419, NL2680, and NL2761). All patients gave written informed consent, which included further use of their data for additional research. The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

Role of the funding source

Parts of this work were supported by the Dutch Arthritis Association (ReumaNederland, project number LLP-9) and by ZonMW (The Netherlands Organisation for Health Research and Development project number 95110008). The sponsors had no role in the study design, collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Acknowledgements

We thank Professor Floris P.J.G. Lafeber for thorough review of the manuscript, input and discussion on the subjects described. We thank Marja Melief for all KIDA analyses.

Parts of this work were supported by the Dutch Arthritis Association (ReumaNederland, project number LLP-9) and by ZonMW (The Netherlands Organisation for Health Research and Development, project number 95110008).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2021.07.013.

References

1. Buckwalter JA, Martin JA. Osteoarthritis. Advanced Drug Delivery Reviews 2006 May 20;58(2):150–67, https://doi.org/10.1016/j.addr.2006.01.006.
2. Vignon E, Piperno M, Le Graverand MPH, Mazzuca SA, Brandt KD, Mathieu P, *et al*. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon schuss views. Arthritis Rheum 2003 Feb 1;48(2):378–84, https://doi.org/10.1002/art.10773.
3. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957 Dec 1;16(4):494–502, https://doi.org/10.1136/ard.16.4.494.
4. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007 Jan 1;15(Suppl 1):1–56, https://doi.org/10.1016/j.joca.2006.06.017.
5. Marijnissen ACA, Vincken KL, Vos PAJM, Saris DBF, Viergever MA, Bijlsma JWJ, *et al*. Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. Osteoarthritis Cartilage 2008 Feb 1;16(2):234–43, https://doi.org/10.1016/J.JOCA.2007.06.009.
6. Vos PAJM, DeGroot J, Huisman AM, Oostveen JCM, Marijnissen ACA, Bijlsma JWJ, *et al*. Skin and urine pentosidine weakly correlate with joint damage in a cohort of patients

with early signs of osteoarthritis (CHECK). Osteoarthritis Cartilage 2010 Oct 1;18(10):1329–36, https://doi.org/10.1016/j.joca.2010.07.006.
7. Intema F, Van Roermund PM, Marijnissen ACA, Cotofana S, Eckstein F, Castelein RM, *et al*. Tissue structure modification in knee osteoarthritis by use of joint distraction: an open 1-year pilot study. Ann Rheum Dis 2011 Aug 1;70(8):1441–6, https://doi.org/10.1136/ard.2010.142364.
8. Holla JFM, Steultjens MPM, van der Leeden M, Roorda LD, Bierma-Zeinstra SMA, den Broeder AA, *et al*. Determinants of range of joint motion in patients with early symptomatic osteoarthritis of the hip and/or knee: an exploratory study in the CHECK cohort. Osteoarthritis Cartilage 2011 Apr 1;19(4):411–9, https://doi.org/10.1016/j.joca.2011.01.013.
9. Kinds MB, Marijnissen ACA, Bierma-Zeinstra SMA, Bijlsma JWJ, Vincken KL, Viergever MA, *et al*. Detection of progression of radiographic joint damage in case of very early osteoarthritis: sensitivity to change of quantitative analysis compared to qualitative grading. Rheumatol Rep 2012 Nov;4(1):27–33, https://doi.org/10.4081/rr.2012.e9.
10. Kinds MB, Vincken KL, Hoppinga TN, Bleys RLAW, Viergever MA, Marijnissen ACA, *et al*. Influence of variation in semiflexed knee positioning during image acquisition on separate quantitative radiographic parameters of osteoarthritis, measured by Knee Images Digital Analysis. Osteoarthritis Cartilage 2012 Sep;20(9):997–1003, https://doi.org/10.1016/j.joca.2012.04.016.
11. Kinds MB, Marijnissen ACA, Vincken KL, Viergever MA, Drossaers-Bakker KW, Bijlsma JWJ, *et al*. Evaluation of separate quantitative radiographic features adds to the prediction of incident radiographic osteoarthritis in individuals with recent onset of knee pain: 5-year follow-up in the CHECK cohort. Osteoarthritis Cartilage 2012 Jun;20(6):548–56, https://doi.org/10.1016/j.joca.2012.02.009.
12. Kinds MB, Vincken KL, Vignon EP, Wolde S Ten, Bijlsma JWJ, Welsing PMJ, *et al*. Radiographic features of knee and hip osteoarthritis represent characteristics of an individual, in addition to severity of osteoarthritis. Scand J Rheumatol 2012;41:141–9, https://doi.org/10.3109/03009742.2011.617311.
13. Wiegant K, van Roermund PM, Intema F, Cotofana S, Eckstein F, Mastbergen SC, *et al*. Sustained clinical and structural benefit after joint distraction in the treatment of severe knee osteoarthritis. Osteoarthritis Cartilage 2013 Nov;21(11):1660–7, https://doi.org/10.1016/j.joca.2013.08.006.
14. Kinds MB, Marijnissen ACA, Viergever MA, Emans PJ, Lafeber FPJG, Welsing PMJ. Identifying phenotypes of knee osteoarthritis by separate quantitative radiographic features may improve patient selection for more targeted treatment. J Rheumatol 2013 Jun 1;40(6):891–902, https://doi.org/10.3899/jrheum.121004.
15. Van Spil WE, Nair SC, Kinds MB, Emans PJ, Hilberdink WKHA, Welsing PMJ, *et al*. Systemic biochemical markers of joint metabolism and inflammation in relation to radiographic parameters and pain of the knee: data from CHECK, a cohort of early-osteoarthritis subjects. Osteoarthritis Cartilage 2015 Jan 1;23(1):48–56, https://doi.org/10.1016/j.joca.2014.09.003.
16. Wesseling J, Bastick AN, Ten Wolde S, Kloppenburg M, Lafeber FPJG, Bierma-Zeinstra SMA, *et al*. Identifying trajectories of pain severity in early symptomatic knee osteoarthritis: a 5-year followup of the cohort hip and cohort knee (CHECK) study. J Rheumatol 2015 Aug 1;42(8):1470–7, https://doi.org/10.3899/jrheum.141036.
17. van der Woude JAD, van Heerwaarden RJ, Spruijt S, Eckstein F, Maschek S, van Roermund PM, *et al*. Six weeks of continuous joint distraction appears sufficient for clinical benefit and cartilaginous tissue repair in the treatment of knee osteoarthritis. Knee 2016 Oct 1;23(5):785–91, https://doi.org/10.1016/j.knee.2016.05.001.
18. van der Woude JAD, Nair SC, Custers RJH, van Laar JM, Kuchuck NO, Lafeber FPJG, *et al*. Knee joint distraction compared to total knee arthroplasty for treatment of end stage osteoarthritis: simulating long-term outcomes and cost-effectiveness. PLoS One 2016 May 12;11(5), e0155524, https://doi.org/10.1371/journal.pone.0155524.
19. van der Woude JAD, Welsing PM, van Roermund PM, Custers RJH, Kuchuk NO, Lafeber FPJGG. Prediction of cartilaginous tissue repair after knee joint distraction. The Knee 2016 Oct;23(5):792–5, https://doi.org/10.1016/j.knee.2016.02.015.
20. Wesseling J, Boers M, Viergever MA, Hilberdink WK, Lafeber FP, Dekker J, *et al*. Cohort profile: cohort hip and cohort knee (CHECK) study. Int J Epidemiol 2016 Feb 1;45(1):36–44, https://doi.org/10.1093/ije/dyu177.
21. van der Woude JAD, Wiegant K, van Heerwaarden RJ, Spruijt S, van Roermund PM, Custers RJH, *et al*. Knee joint distraction compared with high tibial osteotomy: a randomized controlled trial. Knee Surg Sports Traumatol Arthrosc 2017;25(3):876–86, https://doi.org/10.1007/s00167-016-4131-0.
22. Van Der Woude JAD, Wiegant K, Van Heerwaarden RJ, Spruijt S, Emans PJ, Mastbergen SC, *et al*. Knee joint distraction compared with total knee arthroplasty a randomised controlled trial. Bone and Joint Journal 2017;99-B(1):51–8, https://doi.org/10.1302/0301-620X.99B1.BJJ-2016-0099.R3.
23. van der Woude JTAD, Wiegant K, van Roermund PM, Intema F, Custers RJH, Eckstein F, *et al*. Five-year follow-up of knee joint distraction: clinical benefit and cartilaginous tissue repair in an open uncontrolled prospective study. Cartilage 2017;8(3):263–71, https://doi.org/10.1177/1947603516665442.
24. Jansen MP, van der Weiden GS, Van Roermund PM, Custers RJH, Mastbergen SC, Lafeber FPJG. Initial tissue repair predicts long-term clinical success of knee joint distraction as treatment for knee osteoarthritis. Osteoarthritis Cartilage 2018;26(12):1604–8, https://doi.org/10.1016/j.joca.2018.08.004.
25. Besselink NJ, Vincken KL, Bartels LW, van Heerwaarden RJ, Concepcion AN, Marijnissen ACA, *et al*. Cartilage quality (dGEMRIC index) following knee joint distraction or high. Tibial Osteotomy. Cartilage. 2018;11(1):19–31, https://doi.org/10.1177/1947603518777578.
26. Jansen MP, Besselink NJ, van Heerwaarden RJ, Custers RJH, Emans PJ, Spruijt S, *et al*. Knee joint distraction compared with high tibial osteotomy and total knee arthroplasty: two-year clinical, radiographic, and biochemical marker outcomes of two randomized controlled trials. Cartilage 2019 Feb 13;12(2):181–91, https://doi.org/10.1177/1947603519828432.
27. van Helvoort EM, van Spil WE, Jansen MP, Welsing PMJ, Kloppenburg M, Loef M, *et al*. Cohort profile: the Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH) study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical markers. BMJ Open 2020 Jul 28;10(7), e035101, https://doi.org/10.1136/bmjopen-2019-035101.
28. Van Den Berg R, Jongbloed EM, Kuchuk NO, Koes BW, Oei EHG, Bierma-Zeinstra SMA, *et al*. Association between self-reported spinal morning stiffness and radiographic evidence of lumbar disk degeneration in participants of the cohort hip and cohort knee (CHECK) study. Phys Ther 2020 Feb 7;100(2):255–67, https://doi.org/10.1093/ptj/pzz170.

29. van Helvoort EM, Hodgins D, Mastbergen SC, Marijnissen AK, Guehring H, Loef M, *et al*. Relationship between motion, using the GaitSmartTM system, and radiographic knee osteoarthritis: an explorative analysis in the IMI-APPROACH cohort. Rheumatology 2020 Dec 25, https://doi.org/10.1093/rheumatology/keaa809.
30. Jansen MP, Maschek S, van Heerwaarden RJ, Mastbergen SC, Wirth W, Lafeber FFPJG, *et al*. Changes in cartilage thickness and denuded bone area after knee joint distraction and high tibial osteotomy - post-hoc analyses of two randomized controlled trials. J Clin Med 2021 Jan 19;10(2):368, https://doi.org/10.3390/jcm10020368.
31. Jansen MP, Mastbergen SC, Watt FE, Willemse EJ, Vincent TL, Spruijt S, et al. Cartilage Repair Activity during Joint-Preserving Treatment May Be Accompanied by Osteophyte Growth. submitted for publication.
32. Jansen MP, Boymans TAEJ, Custers RJH, Van Geenen RCI, Van Heerwaarden RJ, Huizinga MR, *et al*. Knee joint distraction as treatment for osteoarthritis results in clinical and structural benefit: a systematic review and meta-analysis of the limited number of studies and patients available. Cartilage 2020 Jul 22, https://doi.org/10.1177/1947603520942945. 194760352094294.
33. Jansen MP, Mastbergen SC, Heerwaarden RJ van, Spruijt S, Empelen MD van, Kester EC, *et al*. Knee joint distraction in regular care for treatment of knee osteoarthritis: a comparison with clinical trial data. PLOS ONE 2020 Jan 22;15(1), e0227975, https://doi.org/10.1371/journal.pone.0227975.
34. Buckland-Wright JC, Ward RJ, Peterfy C, Mojcik CF, Leff RL. Reproducibility of the semiflexed (metatarsophalangeal) radiographic knee position and automated measurements of medial tibiofemoral joint space width in a multicenter clinical trial of knee osteoarthritis. J Rheumatol 2004 Aug;31(8):1588–97.
35. Buckland-Wright JC, Wolfe F, Ward RJ, Flowers N, Hayne C. Substantial superiority of semiflexed (MTP) views in knee osteoarthritis: a comparative radiographic study, without fluoroscopy, of standing extended, semiflexed (MTP), and schuss views. J Rheumatol 1999 Dec;26(12):2664–74.
36. Martin Bland J, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb 8;327(8476):307–10, https://doi.org/10.1016/S0140-6736(86)90837-8.
37. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 2016 Jun;15(2):155, https://doi.org/10.1016/J.JCM.2016.02.012.
38. De Vet HCW, Bouter LM, Dick Bezemer P, Beurskens AJHM. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001;17:479–87.
39. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan 1;60(1):34–42, https://doi.org/10.1016/j.jclinepi.2006.03.012.