Lessons learnt about two-year follow-up in hand osteoarthritis: the HOSTAS study.

Damman, W.

9

SENSITIVITY-TO-CHANGE AND VALIDITY OF SEMI-AUTOMATIC JOINT SPACE WIDTH MEASUREMENTS IN HAND OSTEOARTHRITIS: A FOLLOW-UP STUDY

Damman W, Kortekaas MC, Stoel BC, van 't Klooster R, Wolterbeek R, Rosendaal FR, Kloppenburg M

Osteoarthritis Cartilage. 2016;24(7):1172-1179

Abstract

Objective

To assess sensitivity-to-change and validity of longitudinal quantitative semi-automatic joint space width (JSW) measurements and to compare this method with semi-quantitative joint space narrowing (JSN) scoring in hand osteoarthritis (OA) patients.

Methods

Baseline and two-year follow-up radiographs of 56 patients with hand OA (mean age 62 years, 86% women) were used. JSN was scored 0-3 using the Osteoarthritis Research Society International atlas and JSW was quantified in millimetres (mm) in the distal interphalangeal, proximal interphalangeal and second to fifth metacarpophalangeal (DIP, PIP and MCP) joints. Sensitivity-to-change was evaluated by calculating Standardized Response Means (SRMs). A change in JSW or JSN above the Smallest Detectable Difference (SDD) defined progression on joint level. To assess construct validity, progressed joints were compared by cross-tabulation and by associating baseline ultrasound variables with progression (using generalized estimating equations, adjusting for age and sex).

Results

The JSW method detected statistically significant mean changes over 2.6 years (-0.027 mm [95% CI -0.01; -0.04], -0.024 mm [-0.01; -0.03] and -0.021 mm [-0.01; -0.03] for DIP, PIP and MCP joints, respectively). Sensitivity-to-change was low (SRMs: 0.174, 0.168 and 0.211, respectively). 9.1% (121/1336) of joints progressed in JSW, but 3.6% (48/1336) widened. Eighty-three (6.2%) joints progressed in JSW only, 36 (2.7%) in JSN only and 37 (2.8%) in both methods. Progression in JSW showed weaker associations with baseline inflammatory ultrasound features than progression in JSN.

Conclusion

Assessment of progression in hand OA defined by JSW measurements is possible, but performs less well than progression defined by JSN scoring. Therefore, the value of JSW measurements in hand OA clinical trials remains questionable.

Introduction

Hand osteoarthritis (OA) is a prevalent phenotype leading to pain, disability and joint destruction, including cartilage loss[1]. The latter is an important outcome measure in monitoring the disease course[2–5]. Since thickness of cartilage and cartilage loss cannot be directly visualized on conventional radiographs, joint space changes are used as a surrogate.

A widely used, recommended and validated visual grading method to assess the width of joint space is the Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) scoring[6–8]. Changes in JSN are scored by comparison of subsequent radiographs taken over time and sensitivity of the method with trained readers to detect changes in JSN is high[9]. Visual grading methods such as this are considered the 'gold standard' to assess joint space. However, these methods are reader-dependent and even if the reader is experienced, assigning grades remains a subjective process in which the number of grades is limited 0 to 3. Furthermore, change in joint space in the finger joints is small[10–12], making cartilage loss in hand OA over short time periods particularly difficult to assess. Therefore, more objective and sensitive methods are preferred[13–16].

Quantitative joint space width (JSW) measurements present an alternative to JSN scoring. A semi-automatic method to quantify JSW in hand joints, which was shown to be highly accurate and reproducible in phantom and human cadaver hand joints[17], was developed by van 't Klooster et al.[18] and is openly available through www.lkeb.nl (software downloads). This method not only allows an objective manner of JSW measurement in a short time frame without requirement of an experienced reader, but also has the ability to assess widened joints, as was found in patients with acromegaly[19]. These advantages could make quantification of JSW, rather than JSN scoring, useful as an outcome measure to assess small decreases in joint space over short time periods.

The semi-automatic JSW method has been demonstrated to be a valid method to measure JSW in a large cross-sectional population of patients with hand OA and controls[15]. In an earlier longitudinal study in patients with rheumatoid arthritis (RA) the performances of five computer-based JSW methods were studied, suggesting that these JSW measurements are more discriminative in assessing changes than observer scoring[20,21]. Although longitudinal quantitative measurements of JSW in hand OA seem promising, sensitivity-to-change and validity of progression were not studied before in this disease. Therefore, we investigated the performance of JSW measurements longitudinally in a two-year follow-up study in patients with hand OA. First, sensitivity-to-change was assessed and second, construct validity of progression was studied by comparing it with the JSN scoring. With

this paper we aim to increase insight into the performance of semi-automatic JSW measurements in hand OA over time and to ascertain the question whether this method could be useful as an outcome measure in clinical trials.

Materials and methods

Study design

The ECHO study is a longitudinal observational study, in which consecutive patients from the rheumatology outpatient clinic of the Leiden University Medical Center (LUMC) were enrolled between May 2008 and January 2010. Follow-up visits were performed between January 2011 and April 2012. Written informed consent was obtained from all patients and the study was approved by the LUMC medical ethics committee (for details see Kortekaas[22]). All patients fulfilled the American College of Rheumatology (ACR) classification criteria for hand OA[23].

Joints under study

On both hands the distal interphalangeal (DIP), the proximal interphalangeal (PIP) and the second to fifth metacarpophalangeal (MCP) joints were assessed. These 24 joints are the hand joints under study.

Radiographic acquisition

Digital hand radiographs (dorsal-volar views) of both hands were obtained at baseline and follow-up. The radiographic protocol uses a film focus distance of 1.20 meter and a tube voltage of 45kV, 250 mA and 5 mAs with 20msec exposure time (type of film cassette Canon Detector CXDI, pixel spacing 100 microns, grayscale resolution 12-bit).

Radiographic scoring

Hand joints were scored for JSN following the OARSI atlas; per joint a grade of 0 to 3 was given[6]. Since MCP joints are not included in the OARSI atlas, MCP joints were scored based on the PIP atlas. Baseline and follow-up radiographs were scored paired in known order by MCK, who was blind for clinical data. Intrareader reliability for JSN was good, based on randomly selected pairs of radiographs from eight patients (14%) with an intraclass coefficient (ICC) of 0.85 (95% confidence interval [CI]: 0.82 to 0.88). Percentage exact agreement for progression between scoring rounds was 90%. For subgroup analysis, erosions were scored following Verbruggen-Veys anatomical phase scoring and erosive

disease was defined as having joints with eroded (E-phase) or remodelled (R-phase) subchondral plates[24].

Joint space width measurements

JSW was measured on single unpaired radiographs by a semi-automatic quantification method[18]. The image analysis software identifies all joints of interest and the corresponding joint margins and subsequently measures the mean JSW in millimetres (mm) within a measurement interval in each joint, which was determined by the width of the respective phalanx. The automatic results of the image analysis were reviewed by the reader (WD) and corrected if needed. Measurement was blinded for clinical data. Sixty-one radiographs were assessed by two independent persons (WD, SdB). The interindividual variation was low, reflected by an ICC of 0.965 (95% CI 0.96 to 0.97) and a root mean square standard deviation (RMS-SD) of 0.0774.

Definition of progression

Progression on joint level after two years was defined as a change in JSW or JSN above the measurement error.

JSW was measured on single unpaired radiographs and therefore the Smallest Detectable Difference (SDD) was used as a cut-off level for progression[25]. To determine the SDD, we used pairs of radiographs of 22 patients with hand OA with 528 hand joints (from another hand OA cohort; median age 64 years, 82% female, 82% fulfilling ACR criteria for hand OA). These radiographs were acquired in the same manner (same protocol, system and pixel spacing) as in our study group and with a maximum of 196 days in between. We assumed no differences in JSW in such short time. Therefore, the differences could be interpreted as measurement error. The SDD was calculated as $1.96*SD/\sqrt{k}$, in which k is the number of readings or raters (k = 1, because we used one difference (= one reading) between baseline and follow-up of one rater)[25]. SD is the standard deviation of the difference in change scores. Progression was defined as a decrease in JSW more than the SDD.

In JSN scoring the SDD was calculated from data that were used for ICC calculation, which were status-scores. The SDD resulted in 0.86 (SD was 0.44), so progression was defined as an increase of ⩾1 grade. Cut-off levels for SDD were determined per joint group in DIP, PIP and MCP joints. For defining widening, the cut-offs were used in the opposite direction.

Ultrasonography

The ultrasound procedure has been described elsewhere[22]. In brief, it was performed by one experienced ultrasonographer (MCK), scoring in the presence of a second ultrasonographer (WYK) using a Toshiba Applio scanner (Toshiba Medical systems, Tustin, California) with a 10–14 MHz linear array transducer. Each joint was scored for two inflammatory features, being Power Doppler Signal (PDS) and synovial thickening, using a semi-quantitative scale: 0 = none, 1 = mild, 2 = moderate and 3 = severe. Reliability was intermediate-good (ICC for PDS 0.62, synovial thickening 0.93).

Statistical analysis

Mean JSW change in mm on joint level was quantified using linear mixed models (LMM), adjusting for age and sex and was reported for DIP, PIP and MCP joint groups separately (n = 444, 444 and 448, respectively). Joints within patients were defined as the subject variable, as 24 single joints in one patient are not independent observations. Assumptions of normality and constant variance of residuals were met. Missing data in JSW measurement as well as JSN scoring because of a positioning problem in one patient were considered completely at random. BMI data were only taken into account in additional analysis, because of missing data in three patients.

Sensitivity-to-change in JSW was evaluated by calculating Standardized Response Means (SRMs) per joint group (same groups as mean JSW change). SRMs, reflecting the variability of the change scores, were calculated by dividing the average difference by the SD of the differences between the paired measurements. The higher the level of variability in relation to mean change, the smaller the SRM. The 95% CI was calculated as SRM ± 1.96 SD, in which SD is $1/\sqrt{n}$[26,27].

No golden standard is available for assessing progression on radiographs (assessment of cartilage volume would be the preferred 'gold standard'). Hence, testing criterion validity was not possible. Therefore, we investigated construct validity by comparing the JSW and the JSN method. We first did this by cross-tabulation of the number of widened, not-changed and progressed joints defined by the two methods (n = 1336 joints in total). Subsequently, we introduced an external standard, assuming that a decrease in joint space is associated with this standard, i.e., the presence of the inflammatory variables PDS and synovial thickening at baseline, as was earlier shown by Kortekaas[22]. We hypothesized that these associations would be as strong as or stronger for JSW than for JSN.

Reliability of scoring was determined using generalizability theory, as was earlier described by Kortekaas[28]. This method is more suitable than traditional ICC analysis

because it estimates the components of variance within each model, taking into account the outcomes on joint level and joints clustered within a patient. The CI for the ICC was determined using a delta-method approach to estimate the variance of the ICC[29,30]. Additionally, interrater reliability for semi-automatic JSW measurements was determined using the RMS-SD: $\mathrm{RMS-SD} = \sqrt{\frac{\sum \mathrm{SD}^2}{\mathrm{N}}}$[31]. The SD in the equation was estimated by repeat measurements of mean JSW for each of the individual joints (n = 1450, 14 were missing).

Associations of progression with PDS and synovial thickening on joint level were studied using binary logistic generalized estimating equations (GEE) to account for the patient effect (joints within a patient were defined as an within-subject variable in the repeated statement). Odds Ratios (ORs with 95% CIs) were estimated, with progression as the outcome and inflammatory ultrasound features as the determinant, while adjusting for age and sex. An exchangeable correlation matrix was used and joints without the ultrasound feature served as reference. Missing data were handled in the same way as in the LMM.

Two sensitivity analyses were performed. First, only joints at risk for progression were taken into account, omitting joints with a baseline JSN-score of 0 (resulting in incident OA (OA development) instead of OA progression) and 3 (cannot further progress), resulting in n = 589 joint left for analysis. In the second sensitivity analysis joints at risk for widening, i.e., with erosive disease (n = 51), were omitted. Data were analysed using SPSS for Windows, V.20.0 (IBM SPSS statistics, New York, USA).

Results

Study population

Baseline and follow-up radiographs, with a mean (SD) follow-up time of 2.6 (0.3) years, were available of fifty-six patients (mean (SD) age 61.6 (8.9) years, 86% women, mean (SD) BMI 27.6 (4.4) kg/m^2). BMI data were missing in three patients.

Eight joints of the left hand of one patient were impossible to score on the follow-up radiograph due to a positioning problem, leaving 1336 joints available for evaluation of progression. Any JSN was seen at baseline in 674 (50%) joints and at follow-up in 687 (51%) joints in 55 patients; one patient showed no JSN at all. 670 (50%) joints had a baseline JSN score of 0, 441 (33%) a score of 1, 152 (11%) of 2 and 81 (6%) of 3. PDS and synovial thickening were seen at baseline in 89 (6.6%) and 98 (7.3%) joints, respectively. One joint of the right hand could not reliably be assessed for synovial thickening.

Table 1. Joint space width (JSW) at baseline and change over 2.6 years in 1344 joints (448 per joint group) in 56 patients with hand osteoarthritis.

	JSW baseline Mean (SD) in mm	JSW change* Mean (95% CI) in mm	JSW change* Percentage of baseline
DIP joints	0.61 (0.27)	-0.027 (-0.01; -0.04)	- 4.4 %
PIP joints	0.79 (0.26)	-0.021 (-0.01; -0.03)	- 2.7%
MCP joints	1.33 (0.29)	-0.024 (-0.01; -0.03)	- 1.8%

*Adjusted for age and sex. Eight joints were not eligible for evaluation so 1336 were assessed. DIP: distal interphalangeal; PIP: proximal interphalangeal; MCP: metacarpophalangeal.

Quantification of joint space width change over time

The mean (SD) JSW on baseline was in DIP joints almost half the magnitude of MCP joints (0.61 [0.27] mm vs 1.33 [0.29] mm), while the mean of PIP joints was in between (0.79 [0.26] mm). In all joint groups a small but statistically significant decrease in JSW between -0.021 mm and -0.027 mm was seen after 2.6 years (Table 1). Additionally adjusting for BMI did not change the results.

When we stratified the results to baseline JSN score, the mean (SD) baseline JSW in joints with a baseline JSN score of 0 was 1.17 (0.33) mm, while the change (SD; % change of baseline) in JSW in this group after 2.6 years was -0.011 (0.09; -0.9%) mm. Corresponding values for the baseline JSN = 1 group were 0.77 (0.24) and -0.031 (0.11; -4.0%); for the JSN = 2 group 0.52 (0.20) and -0.079 (0.18; -15.2%) and for JSN = 3 group 0.21 (0.22) and 0.010 (0.30; 4.8%).

In Figure 1 these results are combined, so the change per joint group per baseline JSN score is depicted. This figure shows that a decrease in JSW was particularly clear in the baseline JSN = 1 and JSN = 2 groups. In the joints with baseline JSN = 0, the mean JSW stayed the same after 2.6 years and in the joints with JSN baseline score 3 it increased.

Table 2. Concordance between progression on joint level in joint space narrowing (JSN) score and joint space width (JSW) measurements, defined as change above the measurement error in 1336 joints of 56 hand osteoarthritis patients.

JSW \ JSN	Widening	No change	Progression	Total
Widening	5	40	3	48
No change	2	1129	36	1167
Progression	1	83	37	121
Total	8	1252	76	1336

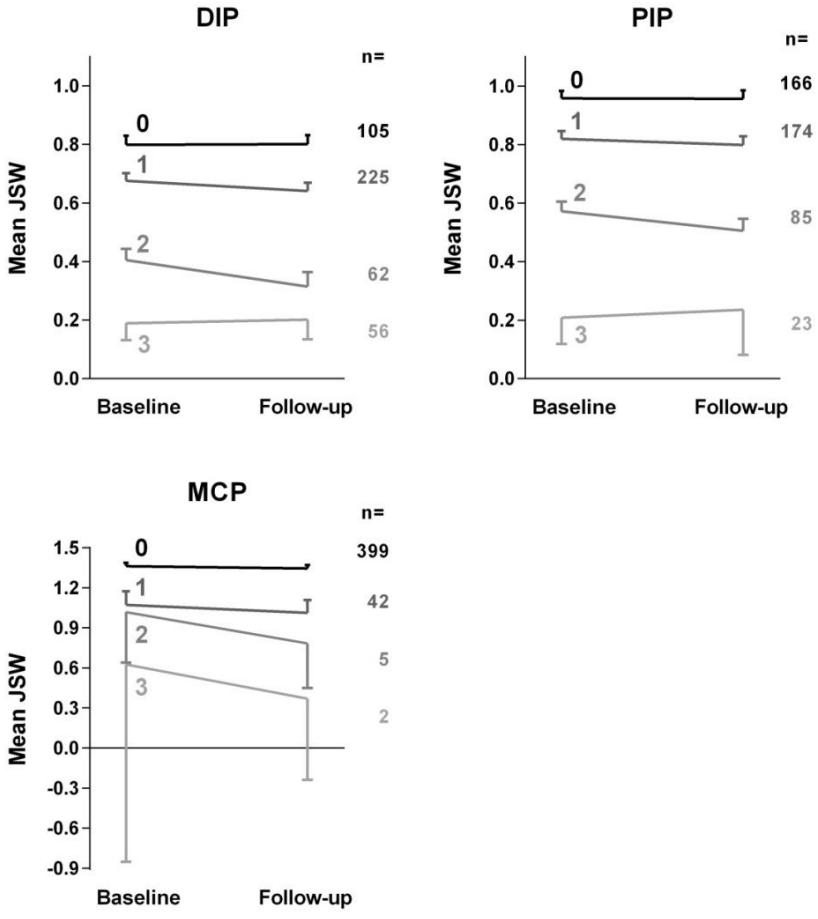

Figure 1. Change in joint space width (JSW) after 2.6 years follow-up for DIP, PIP and MCP joint groups, stratified by joint space narrowing baseline score; 0, 1, 2 or 3. Error bars: 95% CI. n: number of joints.

Progression in JSW and JSN

Progression was defined as a decrease in JSW more than the SDD, resulting in a cut-off of 0.163 mm for DIP joints, 0.109 mm for PIP joints and 0.224 mm for MCP joints, while the SDD for JSN was 1 grade increase. Based on the SDD, 121 joints (9.1%) progressed according to JSW measurements, whereas 76 (5.7%) progressed according to JSN scoring (Table 2).

We hypothesized that the JSW method would be more sensitive, so we expected to find joints classified as 'progressed' with the JSW method, while the JSN method classifies the

same joints as 'no change', which was the situation in 83 joints in our study (69% of 121 joints progressed in JSW).

Following the same hypothesis, some conflicting results were found. Only half (n = 37) of the 76 joints that progressed according to JSN, were also classified as progressed according to the JSW method. In the other joints no change (n = 36) was seen and three were even widened, while we expected joints that progressed in JSN should at least also be classified by JSW as being progressed.

Widening after 2.6 years was seen in more joints (n = 48, 3.6%) in JSW than in JSN (n = 8, 0.6%) measurements (Table 2).

Of joints with increased JSN score after two years, 84% (64/76) also showed a decrease in JSW. This reduction in JSW was significantly greater for joints with JSN progression than for joints with no JSN progression: −0.170 (SD 0.22) mm, n = 76 vs −0.015 (0.11) mm, n = 1260, p<0.000, respectively.

Sensitivity to change of joint space width measurements

Sensitivity-to-change in JSW was evaluated by calculating SRMs per joint group and resulted in low sensitivity-to-change, ranged from 0.168 to 0.211 (Table 3). Sensitivity analysis investigating joints at risk for progression, so only joints with baseline JSN = 1 or JSN = 2, improved the SRMs, ranging from 0.285 to 0.561. Omitting erosive joints did not improve the SRMs, except for a small improvement in DIPs.

Construct validity of radiological progression

Validity of progression defined by the two methods was assessed by associating progression on joint level with inflammatory ultrasound features. Positive associations were found between PDS and synovial thickening at baseline and progression of both JSN and JSW, although these associations were weaker for JSW than for JSN (e.g., PDS grade 1 JSN vs JSW OR 3.5 vs 1.9 and PDS grade 2/3 OR 8.3 vs 5.8) (Table 4). Moreover, synovial thickening was not dose-dependently associated with JSW progression (ORs grades 1, 2 and 3; 4.9, 3.4 and 5.5, respectively). When we additionally adjusted for BMI, similar results were found.

With sensitivity analysis we only analysed joints with baseline JSN score of 1 or 2 (Table 4). Compared with the analysis with all joints, now the association between synovial thickening and JSW progression is dose-dependent (ORs grades 1, 2 and 3; 5.7, 7.5 and 10.8 respectively) and for grades 1 and 2 stronger than with JSN progression (ORs JSN vs

JSW grade 1: 4.3 vs 5.7 and grade 2: 6.2 vs 7.5). However, now PDS grade 1 is not clearly associated anymore with JSW progression (OR [95% CI] 1.6 [0.7; 3.5]).

Table 3. Standardized Response Means (SRMs) and sensitivity analysis when omitting erosive joints (n = 51) or joints with joint space narrowing (JSN) baseline score of 0 or 3 (n = 747) from the analysis.

	SRM		
	All joints n = 1336	Without erosive joints n = 1285	Joints with JSN score 1 or 2 n = 589
DIP joints	0.17 (0.08; 0.27)	0.23 (0.13; 0.33)	0.35 (0.24; 0.47)
PIP joints	0.17 (0.07; 0.26)	0.17 (0.07; 0.26)	0.29 (0.16; 0.41)
MCP joints	0.21 (0.12; 0.30)	0.21 (0.12; 0.30)	0.56 (0.27; 0.85)

DIP: distal interphalangeal; PIP: proximal interphalangeal MCP: metacarpophalangeal.

Table 4. Association of progression in joint space narrowing (JSN) or joint space width (JSW) with power doppler signal (PDS) and synovial thickening on ultrasound in patients with hand osteoarthritis in the ECHO study.

	JSN			JSW		
	Prog	No prog	OR (95% CI)*	Prog	No prog	OR (95% CI)*
All joints, n = 1336						
Synovial thickening						
Grade 0	58	1179	1	94	1143	1
Grade 1	8	44	4.3 (1.7; 10.8)	15	37	4.9 (2.2; 10.8)
Grade 2	6	28	4.7 (1.8; 12.7)	8	26	3.4 (1.5; 7.8)
Grade 3	4	8	9.8 (3.0; 31.7)	4	8	5.5 (1.3; 22.4)
PDS						
Grade 0	58	1189	1	101	1146	1
Grade 1	10	54	3.5 (1.7; 7.0)	11	53	1.9 (1.1; 3.4)
Grade 2+3	8	17	8.3 (3.4; 20.2)	9	16	5.8 (2.2; 15.4)
JSN score of 1 or 2, n = 589						
Synovial thickening						
Grade 0	40	494	1	55	479	1
Grade 1	8	27	4.3 (1.6; 11.5)	12	23	5.7 (2.4; 13.6)
Grade 2	4	9	6.2 (1.8; 21.6)	6	7	7.5 (1.8; 31.3)
Grade 3	4	3	16.0 (3.9; 65.7)	4	3	10.8 (2.1; 55.6)
PDS						
Grade 0	40	499	1	62	477	1
Grade 1	9	28	3.6 (1.8; 7.2)	7	30	1.6 (0.7; 3.5)
Grade 2+3	7	6	13.4 (5.2; 34.7)	8	5	10.5 (3.1; 35.9)

*adjusted for age and sex. Prog: progression; OR: Odds Ratio.

Discussion

This is the first longitudinal study assessing sensitivity-to-change and validity of semi-automatic JSW measurements in hand joints of patients with hand OA. We showed that the JSW method is able to detect small mean changes in mm over 2.6 years in DIP, PIP and MCP joints groups. Sensitivity-to-change, reflected by SRMs per joint group, was low, but improved when only taking joints at risk for progression into account. Many joints (9.1%) passed the threshold for progression as defined by the SDD, but many joints (3.6%) also widened. When evaluating construct validity, both progression in JSW and progression defined by JSN showed significant positive associations with baseline inflammatory ultrasound features. Unfortunately, these associations were weaker for JSW than for JSN, in both complete and sensitivity analyses.

Other automated measurements were done mostly in patients with RA in the PIP and MCP joints and reported larger baseline cross-sectional JSW than we found for these joints groups[32–34]. Also in the study by Kwok et al. in patients with hand OA, who used the same software, a somewhat larger baseline mean JSW was measured[15]. Differences could be due to the exclusion of severely affected joints or to differences in the study population (e.g., a less severely affected hand OA population[15]), in the software, in the films (digitized analogue radiographs[32–34]) and to another underlying disease (RA)[32–34]. Longitudinally, several studies in RA patients, but unfortunately not in patients with OA, were performed[14,33]. For example, Angwin et al. studied two-year changes in computerized JSW measurements of the PIP and MCP joints, reporting good construct validity but larger baseline JSW with larger change than we did[14]. Again, differences could be due to difference in diseases, distributions of age and sex, measured finger joints and measurement failure rate between studies.

Changes in JSW over time can occur for two reasons; disease progression or measurement error. To account for the latter, we chose to use the SDD as a cut-off for progression. We calculated this using two radiographs of one patient. With two radiographs, the SDD reflects the day-to-day variability of hand positioning, radiographic protocol execution and the scoring system. These variations add to a larger measurement error, but approximate reality the best. Nevertheless, the SDD we found for the different joint groups was between 0.109 mm and 0.224 mm and corresponding to SDDs for PIP and MCP joints found in other studies[32,35], supporting the validity of our cut-off. However, the SDD does not account for long-term measurement error due to disease progression, like positioning problems because of increased flexure of the fingers. A study of Angwin et al. showed the relevance of position in hands and reported that with increasing flexure, JSW tended to increase in MCP joints and decrease in PIP joints[35]. The structure of DIP joints is

more comparable to PIP than to MCP joints, so the effect of flexure in DIPs should also be comparable to PIPs.

We found discordant results in classification of progression with the quantitative JSW method and the semi-quantitative JSN method, which could reflect differences in what the methods measure. The JSN scoring depends on the smallest point in joint space, whereas JSW measurements quantify the mean JSW in a predefined interval. For example: a joint with JSN grade 3 (no joints space left), could have a mean JSW more than zero. Furthermore, there was a difference in reading method; radiographs were scored paired in known order for JSN and measured unpaired for JSW. Although the first method is preferable because it is more precise while scoring JSN[36–38], this is not feasible but also not relevant for JSW measurements, as they are semi-automatic.

The amount of widening in JSW (3.6% vs 0.6% in JSN) was more than we hypothesized. Although we argued before that measurement of widening is an advantage of the JSW method, we do not believe this applies to patients with hand OA. In the subjective visual JSN method, due to expectation of the reader that widening is not the course of the disease process in OA, widening could be underestimated. The automated JSW measurement is more objective, but cannot adjust for positioning problems or pseudo-widening as seen in erosive disease like a reader can. Hence, from our data we are unable to conclude what the reason was for widening. Real widening of joint space in hand OA, i.e., thickening of cartilage over time, might have occurred, but was to our knowledge never described.

The most important limitation of our study was the differences in reading methods for JSN and JSW, but several other limitations also apply. Firstly, the radiographs were made in daily clinical practice, consequently protocol variations in acquisition happened, like variations in film focus distance. However, with an SDD also based on this protocol we took this into account. Secondly, an SDD is reader and population dependent and was determined on patients with less severe hand OA. Therefore, this did not completely reflect the measurement errors in our population and could have led to misclassification of progressed joints[39]. Moreover, by dichotomization using and SDD some information may have been lost. Thirdly, we had a relatively small population with severe hand OA, which requires more user interaction in the JSW method, making it prone to measurement errors. In the sensitivity analysis leaving out erosive joints we tried to decrease severity of OA and possible pseudo-widening, but it did not improve the SRM. Probably the SRM remained low because of joints that did not show erosive disease, but did require more user interaction, like joints with high Kellgren-Lawrence score[40]. Finally, semi-automatic measurement methods are dependent on the edge detection algorithm, measurement

region definition and acquisition technique, all of which may affect performance. However, we think that, especially for freely available software like we used, it is important to assess sensitivity-to-change and validity for such methods.

We assessed the performance of semi-automatic JSW measurements over time and compared this with JSN visual grading. Our findings indicate that the JSW method is able to detect change, but, especially in a severe hand OA population, results should be interpreted with caution. Furthermore, the JSW method classifies other joints with progression and shows weaker associations with baseline inflammatory features than the JSN method does. However, JSW measurements could be useful to detect subtle changes in early disease. Joint margins are better defined in early OA, requiring less user interaction and the fingers are not flexed and no erosive disease is present, leading to less measurement error. We found that the variation in JSW in the group with normal JSN was the largest (SD 0.33), but the semi-quantitative JSN-method is not able to differentiate within this group. The JSW method could make it possible to measure a decrease in JSW in early disease, warranting research to explore this hypothesis.

References

1. Kloppenburg M, Kwok W-Y. Hand osteoarthritis--a heterogeneous disorder. *Nat Rev Rheumatol*. 2012;8(1):22-31.
2. Kloppenburg M, Bøyesen P, Smeets W, et al. Report from the OMERACT Hand Osteoarthritis Special Interest Group: Advances and Future Research Priorities. *J Rheumatol*. 2014;41(4):810-818.
3. Bijsterbosch J, Watt I, Meulenbelt I, Rosendaal FR, Huizinga TWJ, Kloppenburg M. Clinical and radiographic disease course of hand osteoarthritis and determinants of outcome after 6 years. *Ann Rheum Dis*. 2011;70(1):68-73.
4. Harris PA, Hart DJ, Dacre JE, Huskisson EC, Spector TD. The progression of radiological hand osteoarthritis over ten years: a clinical follow-up study. *Osteoarthritis Cartilage*. 1994;2(4):247-252.
5. Botha-Scheepers S, Riyazi N, Watt I, et al. Progression of hand osteoarthritis over 2 years: a clinical and radiological follow-up study. *Ann Rheum Dis*. 2009;68(8):1260-1264.
6. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage*. 2007;15, Supplement 1:A1-A56.
7. Visser AW, Bøyesen P, Haugen IK, et al. Radiographic scoring methods in hand osteoarthritis – a systematic literature search and descriptive review. *Osteoarthritis Cartilage*. 2014;22(10):1710-1723.
8. Hunter DJ, Arden N, Cicuttini F, et al. OARSI Clinical Trials Recommendations: Hand imaging in clinical trials in osteoarthritis. *Osteoarthritis Cartilage*. 2015;23(5):732-746.
9. Bijsterbosch J, Haugen IK, Malines C, et al. Reliability, sensitivity to change and feasibility of three radiographic scoring methods for hand osteoarthritis. *Ann Rheum Dis*. 2011;70(8):1465-1467.
10. Pfeil A, Böttcher J, Schäfer ML, et al. Normative Reference Values of Joint Space Width Estimated by Computer-aided Joint Space Analysis (CAJSA): The Distal Interphalangeal Joint. *J Digit Imaging*. 2008;21(1):104-112.
11. Pfeil A, Böttcher J, Seidl BE, et al. Computer-Aided Joint Space Analysis (CAJSA) of the Proximal-Interphalangeal Joint—Normative Age-Related and Gender Specific Data. *Acad Radiol*. 2007;14(5):594-602.
12. Pfeil A, Böttcher J, Seidl BE, et al. Computer-aided joint space analysis of the metacarpal-phalangeal and proximal-interphalangeal finger joint: normative age-related and gender-specific data. *Skeletal Radiol*. 2007;36(9):853-864.
13. Duryea J, Neumann G, Niu J, et al. Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: Analysis of longitudinal data from the osteoarthritis initiative. *Arthritis Care Res*. 2010;62(7):932-937.
14. Angwin J, Lloyd A, Heald G, Nepom G, Binks M, James MF. Radiographic hand joint space width assessed by computer is a sensitive measure of change in early rheumatoid arthritis. *J Rheumatol*. 2004;31(6):1050-1061.
15. Kwok WY, Bijsterbosch J, Malm SH, et al. Validity of joint space width measurements in hand osteoarthritis. *Osteoarthritis Cartilage*. 2011;19(11):1349-1355.
16. Kinds MB, Marijnissen ACA, Bijlsma JWJ, Boers M, Lafeber FPJG, Welsing PMJ. Quantitative radiographic features of early knee osteoarthritis: development over 5 years and relationship with symptoms in the CHECK cohort. *J Rheumatol*. 2013;40(1):58-65.
17. Huétink K, Klooster R van 't, Kaptein BL, et al. Automatic radiographic quantification of hand osteoarthritis; accuracy and sensitivity to change in joint space width in a phantom and cadaver study. *Skeletal Radiol*. 2012;41(1):41-49.

18. van 't Klooster R, Hendriks EA, Watt I, Kloppenburg M, Reiber JHC, Stoel BC. Automatic quantification of osteoarthritis in hand radiographs: validation of a new method to measure joint space width. *Osteoarthritis Cartilage*. 2008;16(1):18-25.
19. Biermasz NR, Klooster R van 't, Wassenaar MJE, et al. Automated image analysis of hand radiographs reveals widened joint spaces in patients with long-term control of acromegaly: relation to disease activity and symptoms. *Eur J Endocrinol*. 2012;166(3):407-413.
20. Sharp JT, Angwin J, Boers M, et al. Multiple Computer-based Methods of Measuring Joint Space Width Can Discriminate Between Treatment Arms in the COBRA Trial — Update of an Ongoing OMERACT Project. *J Rheumatol*. 2009;36(8):1825-1828.
21. Lukas C, Sharp JT, Angwin J, et al. Automated Measurement of Joint Space Width in Small Joints of Patients with Rheumatoid Arthritis. *J Rheumatol*. 2008;35(7):1288-1293.
22. Kortekaas MC, Kwok W-Y, Reijnierse M, Kloppenburg M. Inflammatory ultrasound features show independent associations with progression of structural damage after over 2 years of follow-up in patients with hand osteoarthritis. *Ann Rheum Dis*. 2015;74(9):1720-1724.
23. Altman R, Alarcon G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum*. 1990;33(11):1601-1610.
24. Verbruggen G, Veys EM. Numerical scoring systems for the anatomic evolution of osteoarthritis of the finger joints. *Arthritis Rheum*. 1996;39(2):308-320.
25. Bruynesteyn K, Boers M, Kostense P, Linden S van der, Heijde D van der. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann Rheum Dis*. 2005;64(2):179-182.
26. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness. *J Clin Epidemiol*. 2000;53(5):459-468.
27. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg*. 1998;7(6):565-572.
28. Kortekaas MC, Kwok W-Y, Reijnierse M, et al. Magnetic Resonance Imaging in Hand Osteoarthritis: Intraobserver Reliability and Criterion Validity for Clinical and Structural Characteristics. *J Rheumatol*. 2015;42(7):1224-1230.
29. Johnson NL, Kotz S, Kemp AW. *Univariate Discrete Distributions*. 1982.
30. Norusis MJ. *SPSS 14.0 Advanced Statistical Procedure Companion*. Upper Sadle River, NJ, USA: Prentice Hall Inc; 2005.
31. Chen Y, Qiang M, Zhang K, Li H, Dai H. A reliable radiographic measurement for evaluation of normal distal tibiofibular syndesmosis: a multi-detector computed tomography study in adults. *J Foot Ankle Res*. 2015;8(1):32.
32. Peloschek P, Langs G, Weber M, et al. An automatic model-based system for joint space measurements on hand radiographs: initial experience. *Radiology*. 2007;245(3):855-862.
33. Finckh A, de Pablo P, Katz JN, et al. Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: A longitudinal study. *Arthritis Rheum*. 2006;54(5):1444-1450.
34. Neumann G, dePablo P, Finckh A, Chibnik LB, Wolfe F, Duryea J. Patient repositioning reproducibility of joint space width measurements on hand radiographs. *Arthritis Care Res*. 2011;63(2):203-207.
35. Angwin J, Heald G, Lloyd A, Howland K, Davy M, James MF. Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis. *J Rheumatol*. 2001;28(8):1825-1836.

36. Botha-Scheepers S, Watt I, Breedveld FC, Kloppenburg M. Reading radiographs in pairs or in chronological order influences radiological progression in osteoarthritis. *Rheumatology*. 2005;44(11):1452-1455.

37. Auleley G-R, Giraudeau B, Dougados M, Ravaud P. Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies. *Ann Rheum Dis*. 2000;59(6):422-427.

38. Tuyl LHD van, Heijde D van der, Knol DL, Boers M. Chronological reading of radiographs in rheumatoid arthritis increases efficiency and does not lead to bias. *Ann Rheum Dis*. 2014;73(2):391-395.

39. 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.

40. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. *Ann Rheum Dis*. 1957;16(4):494-502.