Radiographic scoring methods in hand osteoarthritis - a systematic literature search and descriptive review

Visser, A.W.; Boyesen, P.; Haugen, I.K.; Schoones, J.W.; Heijde, D.M. van der; Rosendaal, F.R.; Kloppenburg, M.

Osteoarthritis and Cartilage 22 (2014) 1710–1723

Osteoarthritis and Cartilage

Radiographic scoring methods in hand osteoarthritis – a systematic literature search and descriptive review

A.W. Visser † *, P. Bøyesen ‡, I.K. Haugen ‡, J.W. Schoones §, D.M. van der Heijde † ‡, F.R. Rosendaal ||, M. Kloppenburg † ||

† *Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands*
‡ *Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway*
§ *Walaeus Library, Leiden University Medical Center, Leiden, The Netherlands*
|| *Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands*

ARTICLE INFO

Article history:
Received 21 January 2014
Accepted 30 May 2014

Keywords:
Osteoarthritis
Hand
Radiography
Systematic review

SUMMARY

Objective: This systematic literature review aimed to evaluate the use of conventional radiography (CR) in hand osteoarthritis (OA) and to assess the metric properties of the different radiographic scoring methods.

Design: Medical literature databases up to November 2013 were systematically reviewed for studies reporting on radiographic scoring of structural damage in hand OA. The use and metric properties of the scoring methods, including discrimination (reliability, sensitivity to change), feasibility and validity, were evaluated.

Results: Of the 48 included studies, 10 provided data on reliability, 11 on sensitivity to change, four on feasibility and 36 on validity of radiographic scoring methods. Thirteen different scoring methods have been used in studies evaluating radiographic hand OA. The number of examined joints differed extensively and the obtained scores were analyzed in various ways. The reliability of the assessed radiographic scoring methods was good for all evaluated scoring methods, for both cross-sectional and longitudinal radiographic scoring. The responsiveness to change was similar for all evaluated scoring methods. There were no major differences in feasibility between the evaluated scoring methods, although the evidence was limited. There was limited knowledge about the validity of radiographic OA findings compared with clinical nodules and deformities, whereas there was better evidence for an association between radiographic findings and symptoms and hand function.

Conclusions: Several radiographic scoring methods are used in hand OA literature. To enhance comparability across studies in hand OA, consensus has to be reached on a preferred scoring method, the examined joints and the used presentation of data.

Introduction

Osteoarthritis (OA) is the most common musculoskeletal disorder, frequently affecting the hands[1,2]. Hand OA is characterized by the formation of bony enlargements and deformities, most frequently occurring in the distal interphalangeal (DIP) joints and first carpometacarpal (CMC1) joints, less often in the proximal interphalangeal (PIP) joints and least prevalent in metacarpaphalangeal (MCP) joints[3]. Currently, no structure modifying treatments are available. To date, few high-quality clinical trials have been performed in hand OA[4,5]. A key problem in the lack of high-quality clinical trials in hand OA is the lack of standardization of outcome measures[4,6]. The Outcome Measures in Rheumatoid Clinical Trials (OMERACT) and Osteoarthritis Research Society International (OARSI) Task Force on Clinical Trials Guidelines defined core domains to describe outcomes in clinical trials. One of these domains for structure modifying trials was imaging.[7–9]

Conventional radiography (CR) is commonly used to assess structural damage in hand OA, as they are widely available and relatively cheap. Radiography allows visualization of osteophytes, joint space narrowing (JSN), subchondral cysts, sclerosis and central erosions.

* Address correspondence and reprint requests to: Visser A.W., Leiden University Medical Center, Department of Rheumatology, C1-R, P.O. Box 9600, 2300 RC Leiden, The Netherlands. Tel: 31-71-5263265; Fax: 31-71-5266752.
E-mail address: a.w.visser@lumc.nl (A.W. Visser).

http://dx.doi.org/10.1016/j.joca.2014.05.026
1063-4584/

Several standardized scoring methods are available such as the Kellgren–Lawrence (KL)[10], Kessler[11] and Kallman grading scales[12], the OARSI scoring atlas[13], the Verbruggen–Veys anatomical phase score[14], and the Gent University scoring system (GUSS)[15]. These scores differ in the joints that are assessed, the type of scores (composite score or individual feature scores), and the total score ranges.

Most scoring methods have been shown to be reliable instruments for the assessment of structural damage in hand OA as well as its change[15–17]. However, a systematic comparison of the different scoring methods that will help to decide on a recommended method has not been performed.

We performed a systematic review to evaluate the use of CR in studies on hand OA and to assess the metric properties of the different radiographic scoring methods[18]. To this end we made use of the OMERACT filter[19], focusing on aspects of discrimination (reliability and sensitivity to change), feasibility and truth (validity) of the radiographic scoring methods available in hand OA.

Methods

Identification of studies

In cooperation with a medical librarian (JWS), a systemic literature search was performed to obtain all manuscripts reporting on any radiographic scoring methods assessing the nature, severity and progression of structural damage in hand OA. Medical literature databases (PubMed, Embase, Web of Science, COCHRANE and CINAHL) were searched up to November 2013, using all variations of the following key words 'hand', 'osteoarthritis', 'radiography', 'reliability', 'validity', 'sensitive' and 'feasibility' (see Supplementary File For Exact Search Strings).

Inclusion and exclusion criteria

First all retrieved titles were screened, subsequently selected abstracts were reviewed and finally full text articles of the remaining references were read by one reviewer (AWV). A random sample of 150 titles was also reviewed by a second reviewer (MK), resulting in a similar selection of titles. In case of uncertainties in the reviewing process by the single reviewer, these were discussed and solved with MK. The metric properties of the studied radiographic scoring methods were evaluated according to four items: reliability, sensitivity to change, feasibility and validity. Inclusion criteria required for studies to evaluate these items differed per item:

- Reliability was evaluated in studies describing the reliability of two or more scoring methods performed on the same radiographs and by the same reader. Both cross-sectional and longitudinal studies were included.
- Sensitivity to change was evaluated in longitudinal studies of at least one year, in which hand OA was assessed by at least two radiographic scoring methods. Studies with a follow-up duration between one and three years using only one radiographic scoring method were also included.
- Feasibility was evaluated in studies describing the feasibility of one or more scoring methods.
- Validity was evaluated in studies comparing a radiographic scoring method with other measurements of structural damage such as magnetic resonance imaging (MRI), computed tomography (CT), ultrasound (US), digital photography, histology or nodes at physical examination. In addition, validity was evaluated in studies comparing radiographic findings to clinical signs such as hand function or symptoms. Both cross-sectional and longitudinal studies were included.

Studies that fulfilled the requirements for at least one of these four items were included in this review.

Animal studies, reviews, abstracts, letters to the editor and studies reporting on musculoskeletal diseases other than hand OA or in languages other than English were excluded.

Data extraction

A standardized form was used to extract information about the following data: (1) study population (population size, setting, age, sex), (2) applied radiographic scoring methods, (3) performance of the scoring (number of readers, consensus/independent reading, (4) assessed joints, (5) level of analyses of obtained scores (joint, joint group or patient level) and used definition of outcome (e.g., summed scores (total or per feature), counts of number of affected joints, dichotomized outcome), (6) results concerning: reliability (intraclass correlation coefficient (ICC), kappa-value, percentage of agreement, smallest detectable change (SDC)), sensitivity to change (percentage of change, amount of change, standardized response mean (SRM)), feasibility (time needed to perform scoring), validity (correlations, associations and measures of agreement between radiographic scores and other measures). From a random number of studies data were also extracted by MK and all extracted results were discussed with MK.

Statistical analyses

Due to the heterogeneity of the studies and the difference in outcome measures that were used it was not possible to perform a meta-analysis. Therefore we chose to perform a descriptive review.

Results

Literature flow

After removing duplicate references, 1873 unique references were identified [Fig. 1]. After reviewing 133 abstracts and 80 full-text articles, 48 articles were included in this review. Of the included studies, 10 fulfilled the inclusion criteria for evaluation of reliability[12,16,17,20–26], 11 for sensitivity to change[14,16,17,24–31], four for feasibility[11,16,17,22], and 36 for validity of radiographic scoring methods.[20–24,32–62]

Evaluation of radiographic scoring methods was the primary aim in 10 of the included studies[11,12,14,16,17,22,26,27,59,60]. The other studies used radiographic scoring to identify prevalence or progression of radiographic OA features ($n = 7$)[20,25,28–30,33,34], or to compare obtained scores with other outcome measures (other imaging methods, clinical outcomes, histology) ($n = 31$).[21,23,24,31,32,35–38,40–58,61–63]

The characteristics of the evaluated or applied radiographic scoring methods (except for non-validated methods) are depicted in Table I.

Study characteristics

The characteristics of the 48 included studies are depicted in Table II. Most studies included more women than men and most of the studied individuals were aged >50 years. As shown in Table II, a wide variety of scoring methods ($n = 13$) was used to assess radiographic (signs of) hand OA. The KL scoring method was used most frequently ($n = 24$), followed by the OARSI scoring method ($n = 18$). Other scoring methods were the Kallman ($n = 9$), individual features following non-validated methods ($n = 7$),

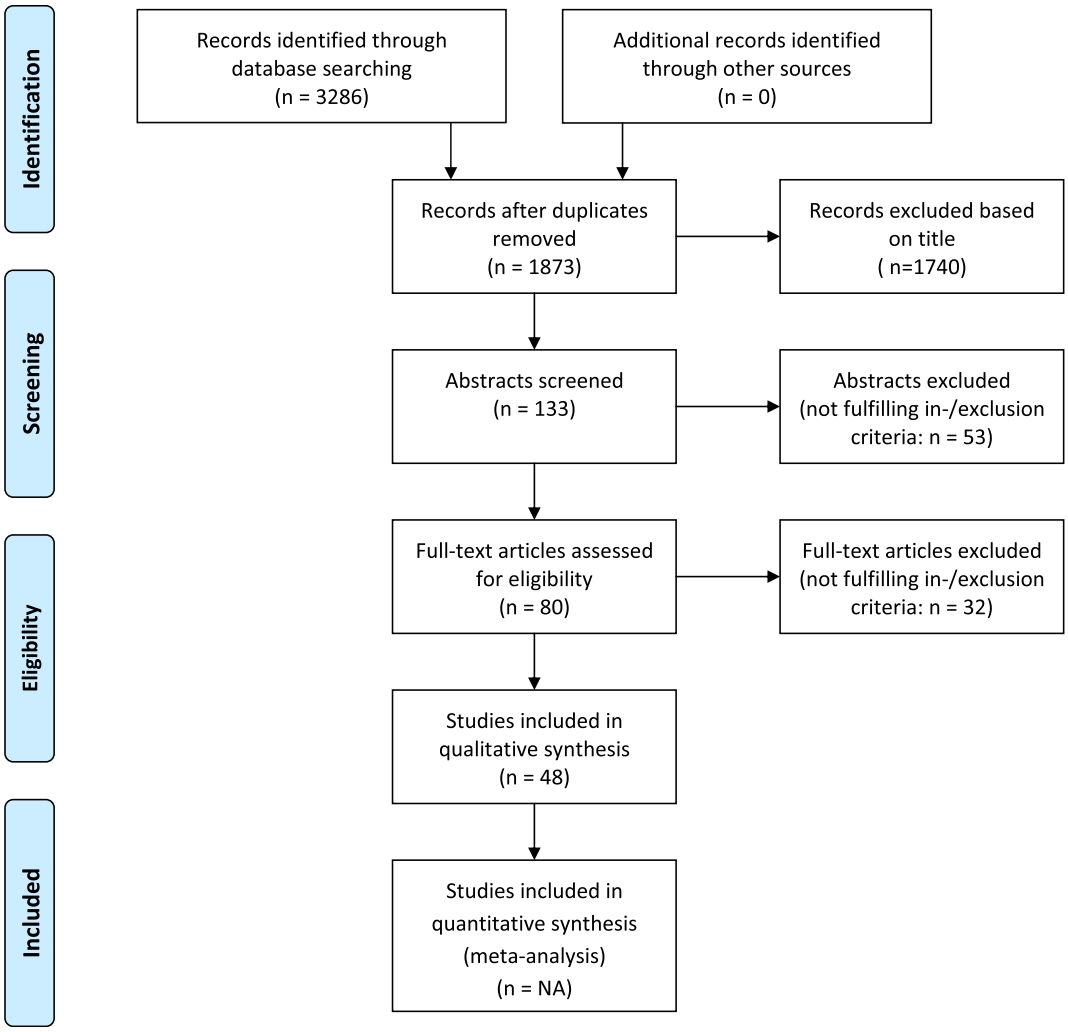

Fig. 1. Overview of literature research.

anatomical phases ($n = 6$), anatomical lesions ($n = 2$) and automatic JSW measurement ($n = 3$). The GUSS, Burnett, Kessler, Lane, Eaton and a non-validated global score were all used in only one study. Although the majority of studies used only one radiographic scoring method, 15 studies used more than one method.

The examined joint groups differed between the studies: DIPs and PIPs were assessed most frequently (in 48 and 46 studies, respectively), followed by the CMC1s ($n = 34$), MCPs ($n = 30$), IP1s ($n = 23$) and the scaphotrapezotrapezoidal (STT) joints ($n = 8$).

The way the analysis of the radiographic scores were executed was quite different across the studies; (1) the score of one joint (the most severely affected) from a joint group, hand or patient[33,36,37,43,46,50], (2) sum score for all joints and features[14,16,17,20–22,24–26,31,34,38,44,45], (3) sum scores per feature[21,22,24,27–29,48], (4) sum scores per joint group[16,24,47,49], (5) mean score per feature[12,30] or per joint[60], (6) scores on joint level (composite score or per feature)[12,20–24,34,35,38,40–44,47,48,51–53,60,61] and (7) presence or absence of radiographic features per joint[21,22,54,55,57,58], joint group[32,38,39,45], or on patient level[52,56].

Discrimination

Reliability

Ten included articles provided data on the reliability of at least two radiographic scoring methods, shown in Table III. The KL scoring method was assessed in seven of these studies[12,16,17,20,21,23,24]. Other assessed scoring methods were the Kallman ($n = 4$)[12,17,20,23], OARSI ($n = 4$)[16,21,22,24], anatomical phases ($n = 4$)[16,17,25,26], anatomical lesions ($n = 1$)[26], GUSS ($n = 1$)[25], global score ($n = 1$)[17], and the semi-automated joint space width (JSW) measurement ($n = 1$).[22]

Eight studies provided cross-sectional data[12,16,17,20–24]. The ICCs as well as kappa values were shown to be reliable for all assessed total scores, and no differences between the scoring methods were observed. The ICCs and kappa values for the individual radiographic features depended on the scored feature; the lowest reliability was reported for the scoring of cysts and the highest for the scoring of erosions and osteophytes.[12,20,21]

In five of the studies readers performed the scoring independently of another reader, providing results on the interreader reliability[12,16,17,21,24]. The interreader ICCs and kappa values were somewhat lower than the intrareader values, especially for the Kallman method and for sclerosis as scored using the OARSI atlas[12,17,24]. Whether readers were from one or different centers did not seem to influence the reliability of the scoring methods.

Six studies provided data on the reliability of change of at least two radiographic scoring methods[12,16,17,24–26]. The reliability of change of KL, OARSI, Kallman, global, anatomical phases and GUSS scores was reported to be good for all methods[12,16,17,24–26]. Bijsterbosch *et al.* compared the SDC of three scoring methods on

Table I
Radiographic scoring methods for hand osteoarthritis

Scoring method	No. of joints	DIP	PIP	IP1	MCP	CMC1	STT	Scored features	Type of score	Range of total score
Anatomical phases[14]	26	+	+	+	+	−	−	Osteophytes, JSN, erosions, sclerosis	Composite score	0–218.4
Anatomical lesions[14]	24	+	+	−	+	−	−	Osteophytes, JSN, cysts	Composite score	Not specified
Burnett[74]	18	+	+	−	−	+	−	Osteophytes, JSN, sclerosis	Individual features	0–126
Eaton[75]	4	−	−	−	−	+	+	Osteophytes, JSN, erosions, cysts, sclerosis, subluxation	Composite score	Not specified
GUSS[15]	18	+	+	+	−	−	−	Osteolytic areas, bone plate resorption, JSN	Composite score	10–300
Kallman[12]	22	+	+	+	−	+	+	Osteophytes, JSN, cysts, sclerosis, deformity, cortical collapse	Individual features	0–208
Kellgren-Lawrence[10]	30	+	+	+	+	+	−	Osteophytes, JSN, sclerosis, alignment	Composite score	0–120
Kessler[11]	18	+	+	−	−	+	−	Osteophytes, JSN, sclerosis	Composite score	0–18
Lane[76]	22	+	+	+	−	+	+	Osteophytes, JSN, erosions/cysts, sclerosis, deformity	Individual features	0–182
OARSI[13]	20	+	+	+	−	+	−	Osteophytes, JSN, erosions/cysts, sclerosis, alignment	Individual features	0–198

Abbreviations: CMC1 = First carpometacarpal joint, DIP = distal interphalangeal joint, IP1 = First interphalangeal joint, MCP = metacarpaphalangeal joint, No. = number, PIP = proximal interphalangeal joint, STT = scaphotrapezotrapezoidal joint.

patient level, showing a small difference in favor of the KL score, followed by the anatomical phases and OARSI scores. Reported SDCs were a little higher over a 6 year interval than over a 2 year interval[16]. Haugen *et al.* assessed reliability of change in KL and OARSI scores, showing a good reliability for the KL score and most of the OARSI features. ICC and kappa values were somewhat lower for change scores than for baseline KL and OARSI scores. Except for change of sclerosis (OARSI), moderate to good reliability was reported for the scoring of change in KL and OARSI scores[24]. Kallman *et al.* evaluated agreement on progression in KL and Kallman scores on joint group level, showing that agreement was more often present in DIP joints than PIP joints and that agreement was lowest on the progression of cysts.[12]

Sensitivity to change

Table IV shows the characteristics of the included studies describing data on sensitivity to change of radiographic scoring methods. Nine studies reported data on short-term follow-up (≤3 years), most of them on patient level[16,17,25–31]. Two studies evaluated change of summed KL, Kallman and anatomical phases scores, of which one study also evaluated the global score[16,17]. Maheu *et al.* reported SRMs over a 1 year interval of the global, KL, Kallman, anatomical phases and OARSI scores; all below 0.50, indicating that the responsiveness to change was small[17]. Bijsterbosch *et al.* detected somewhat more progression over a 2 year interval when scored following the KL or anatomical phases score as compared with the OARSI atlas[16]. The anatomical phases score was evaluated in two other studies[25,26], one of these studies (a randomized controlled trial (RCT)) also assessed change of GUSS. Progression over a 1 year interval was detected by both scoring methods, although no difference between treatment and placebo group was observed.[25]

Five studies reported follow-up data of only one scoring method[27–31]. Botha-Scheepers *et al.* reported change of JSN and osteophytes as scored following the OARSI atlas over a 2 year interval[27–29]. Scoring of these features tended to be more sensitive to change when scoring radiographs in chronological order as compared with paired reading[27]. Buckland–Wright *et al.* evaluated stereoscopic measurement of individual OA features during a 1.5 year interval, reporting change of most features[64]. Olejárová *et al.* evaluated change of hand OA over a 2 year interval using the Kallman scoring method, reporting no significant difference in total score.[31]

In the three studies investigating long term follow-up data (>3 years), change in KL ($n = 2$), OARSI ($n = 2$), anatomical phases ($n = 2$) and anatomical lesions ($n = 1$) score was evaluated[12,14,16,24]. Studies with a longer follow-up duration detected higher occurrence of progression of OA features as well as higher mean radiographic change scores.[16]

Feasibility

Four studies reported data regarding feasibility of radiographic scoring methods (Table V)[11,16,17,22]. The KL, anatomical phases and Kallman scoring methods were assessed in two studies[16,17]. The OARSI, Kessler and Lane scoring methods, as well as a non-validated global score and semi-automated JSW measurement, were all examined in only one study.[11,16,17,22]

The mean time to perform scoring ranged from 1.5 to 10–15 min per hand radiograph. The KL, anatomical phases and Kessler scoring methods seemed to be least time consuming while scoring according Kallman, Lane and the OARSI atlas needed more time to perform[11,16,17]. However, the time needed to perform the scoring differed per study[11,16,17]. Bijsterbosch *et al.* showed that the performance time increased in patients with higher levels of structural abnormalities; 1 min increment in performance time was associated with 3.9 points in KL score (95% confidence interval (CI) 1.0, 6.8), 8.0 (5.3, 10.7) points in OARSI score, and 21.1 (12.9, 29.2) points in the anatomical phases scoring method.[16]

Validity

The 36 studies providing data regarding validity of radiographic scoring methods are listed in Table VI. Analyses on individual joint level were performed in 18 of these studies, and analyses on joint group or patient level were performed in 13 and 14 studies, respectively.

Thirteen studies focused on structural findings at physical examination in comparison to radiographic OA findings[20,22,33–42]. Four studies presented correlation coefficients and kappa values, reporting that nodes at physical examination were weakly to moderately associated with radiographic hand OA[34,35,37,38]. The lowest agreement was reported in a study on clinical Heberden nodes and radiographic DIP osteophytes scored following the Burnett scoring method, performed on joint level ($k = 0.36$)[35]. The highest correlation was reported in a study examining a clinical score consisting of nodes and deformity and the radiographic KL score, analyzed on joint group level (males $r = 0.47$, females $r = 0.66$).[38]

Two studies reported the association between two radiographic scoring methods and clinical nodes, both analyzed on a joint level[20,41]. Addimanda *et al.*, examining KL and Kallman scores, reported the erosion and osteophyte features of the Kallman method to be associated most strongly with nodes (OR 7.4 and 3.2

Table II
Overview of included studies ($n = 48$)

First author, year of publication	Source population, **no. of patients (% women)**, mean age (years)	Scoring methods	Joints investigated	Analysis of radiographic scores
Addimanda, 2012[20]	Secondary care (50% erosive OA), **446 (93)**, 68	KL	DIP, PIP, CMC1	Score per joint, summed total
		Kallman	DIP, PIP, CMC1	Score per joint per feature, summed per joint, summed total
Bagge, 1991[33]	General population, **217 (66)**, 82	KL	DIP, PIP, IP, MCP, CMC1	Score per joint group (most affected joint)
Bijsterbosch, 2011[16]	Familial polyarticular OA (GARP), **90 (78)**, 60	KL	DIP, PIP, IP1, MCP, CMC1	Summed per joint group, summed total
		OARSI	DIP, PIP, IP1, CMC1	Summed per joint group, summed total
		Anatomical phases	DIP, PIP, IP1, MCP	Summed per joint group, summed total
Botha-Scheepers, 2005[27]	Familial polyarticular OA (GARP), **20 (90)**, median age 62	OARSI	DIP, PIP, IP1, MCP, CMC1, STT	Summed total per feature
Botha-Scheepers, 2007[29]	Familial polyarticular OA (GARP), **193 (80)**, 60	OARSI	DIP, PIP, IP1, MCP, CMC1, STT	Summed total per feature
Botha-Scheepers, 2009[28]	Familial polyarticular OA (GARP), **172 (79)**, 61	OARSI	DIP, PIP, IP1, CMC1	Summed total per feature
Buckland –Wright,1990[30]	Unclear (radiographic OA patients), **32 (91)**, 62	Stereoscopic measurement	DIP, PIP, MCP	Mean score total per feature, mean score per joint group per feature
Caspi, 2001[34]	Secondary care (geriatric patients), **253 (68)**, 79	Modified OARSI	DIP, PIP, IP1, MCP, CMC1	Score per joint, summed total
Ceceli, 2012[62]	Secondary care, **60 (100)**, 59	Kallman	Not specified	Summed per hand
Cicuttini, 1998[35]	General population (twin study), **660 (100)**, 56	Burnett	DIP	Score per joint
		Kallman	PIP, CMC1	Score per joint
Dahaghin, 2004[43]	General population (Rotterdam study), **3906 (58)**, 67	Modified KL	DIP, PIP, MCP, CMC1, STT	Score per joint, score per joint group, score per patient (most affected joint)
Ding, 2007[44]	Finnish dentists/teachers, **543 (100)**, range 45 –63	KL	DIP, PIP, IP1, MCP	Score per joint, no. of joints scored ≥2, summed total
Dominick, 2005[45]	Familial OA (Genetics of Generalized Osteoarthritis (GOGO) study), **700 (80)**, 69	KL	DIP, PIP, IP1, MCP, CMC1, STT	Present/absent of score ≥2 per joint group, summed total
Drape, 1996[32]	Secondary care (mucoid cyst), **23 (61)**, 63	Osteophytes, JSN (NVM)	DIP	Present/absent per joint group per feature
El-Sherif, 2008[46]	Secondary care, **40 (100)**, 57	KL	DIP, PIP, IP1, MCP, CMC1	Score per patient (most affected joint)
Grainger, 2007[54]	Secondary care, **15 (93)**, 59	Erosions (NVM)	DIP, PIP	Present/absent per joint
Hart, 1991[36]	Primary/secondary care (non-joint related problems), **541 (100)**, 54	KL	DIP, PIP, CMC1	Score per joint group (most affected joint)
Hart, 1994[37]	Primary care, **976 (100)**, age range 45–65	KL	DIP, PIP, CMC1	Score per joint group (most affected joint)
Haugen, 2012[21]	Secondary care (Oslo hand OA cohort), **106 (92)**, 69	KL	DIP, PIP, IP1, MCP, CMC1	Score per joint, summed total
		OARSI	DIP, PIP, IP1, MCP, CMC1	Score per joint per feature, summed total per feature
		Marginal erosions (NVM)	DIP, PIP, IP1, MCP, CMC1	Present/absent per joint
Haugen, 2013[24]	Secondary care (Oslo hand OA cohort), **190 (91)**, 62 (longitudinal analysis: **99 (92)**, 61)	KL	DIP, PIP, IP1, MCP, CMC1	Score per joint, summed per joint group, summed total
		OARSI	DIP, PIP, IP1, MCP, CMC1	Score per joint per feature, summed total per feature
Huetink, 2012[59]	22 phantom joints, 22 human cadaver joints	Automatic JSN quantification	DIP, PIP, MCP	Millimeter (mm) per joint
Iagnocco, 2005[56]	Secondary care (inflam-matory OA), **110 (100)**, 67	Classical/erosive OA (NVM)	DIP, PIP	Present/absent per patient
Jones, 2001[47]	Secondary care, **522 (67)**, 56	OARSI	DIP, CMC1	Score per joint per feature, summed per joint group
Jonsson, 2012[38]	General population (AGES-Reykjavik study), **381 (58)**, 76	KL	DIP, PIP, CMC1	Score per joint, present/absent of score ≥2 per joint group, summed total
Kallman, 1989[12]	General population (BLSA), **50 (0)**, 68	KL	DIP, PIP, IP1, CMC1	Score per joint, score per joint group, mean score total
		Kallman	DIP, PIP, IP1, CMC1, STT	Score per joint per feature, score per joint group per feature, mean score total per feature
Keen, 2008[57]	Secondary care, **37 (84)**, 57	OARSI	DIP, PIP, MCP, CMC1	Present/absent per joint per feature
Kessler, 2000[11]	Advanced hip/knee OA patients (Ulm OA study) **50**, range 51–79	Kessler	DIP, PIP, CMC1	No. of affected joints per joint group
		Kallman	DIP, PIP, CMC1	Not specified
		Lane	DIP, PIP, CMC1	Not specified
Kortekaas, 2011[48]	Secondary care, **55 (47)**, 61	OARSI	DIP, PIP, IP1, CMC1	Score per joint per feature, summed total per feature
Kwok, 2011[22]	Familial polyarticular OA (GARP), **235 (83)**, 65, and 471 controls	OARSI	DIP, PIP, MCP	Score per joint per feature, summed total per feature
		Anatomical phases	DIP, PIP	Present/absent per joint
		Semi-automated measured JSW	DIP, PIP, MCP	Score per joint, summed total
Lee, 2012[49]	General population (KLoSHA), **378 (48)**, 75	KL	DIP, PIP, IP1, MCP, CMC1	Summed per finger
Maheu, 2007[17]	Secondary care, **105 (93)**, 61	KL	DIP, PIP, MCP, CMC1	Summed total
		Kallman	DIP, PIP, MCP, CMC1,STT	Summed total
		Global score	DIP, PIP, MCP, CMC1,	Summed total
		Anatomical phases	STT DIP, PIP, MCP	Summed total
Mancarella, 2010[23]	Secondary care, **35 (94)**, 66	KL	DIP, PIP, MCP	Score per joint
		Kallman	DIP, PIP, MCP	Score per joint
Marshall, 2009[39]	Primary care (hand pain), **592 (62)**, 64	KL	DIP, PIP, IP1, MCP, CMC1, STT	Present/absent of score ≥2 per joint group
Mathiessen, 2012[40]	Secondary care (Oslo hand OA cohort), **127 (91)**, 69	OARSI	DIP, PIP, IP1, MCP	Score per joint per feature

Table II (*continued*)

First author, year of publication	Source population, **no. of patients (% women)**, mean age (years)	Scoring methods	Joints investigated	Analysis of radiographic scores
Olejárová, 2000[31]	Secondary care, erosive OA: **28 (93)**, 68; non-erosive OA: **24 (83)**, 65	Kallman	DIP, PIP, IP1, MCP, CMC1	Summed total
Ozkan, 2007[50]	Secondary care, **100 (87)**, 69	KL	DIP, PIP, MCP, CMC1	Score per patient (most affected joint)
Rees, 2012[41]	Secondary care (Genetics of Osteoarthritis and Lifestyle (GOAL) study participants with ≥1 node), **1,939 (54)**, 68	KL	DIP, PIP, IP1, CMC1	Score per joint
		OARSI	DIP, PIP, IP1, CMC1	Score per joint per feature
Saltzherr, 2013[61]	Secondary care, **30 (70)**, median age 57	Eaton	CMC1, STT	Score per joint, score per joint per feature
Sonne-Holm, 2006[51]	General population (Copenhagen city hearth study), **3,355 (61)**,age>20	Modified KL	CMC1	Score per joint, score per joint per feature
Stern, 2004[42]	Primary and secondary care (Investigation of Nodal Osteoarthritis to Detect an Association with Loci encoding IL-1 (I-NODAL) study), **71 (80)**, 67	KL	DIP, PIP, IP1, CMC1	Score per joint
Sunk, 2012[53]	Post mortem IP joints, **40 (44)**, median age 66	KL	DIP, PIP	Score per joint
		OARSI	DIP, PIP	Score per joint per feature
Verbruggen, 1996[14]	Unclear (radiographic OA), **46 (96)**, 57	Anatomical phases	DIP, PIP, MCP	Summed total
		Anatomical lesions	DIP, PIP, MCP	Summed total
Verbruggen, 2002[26]	Unclear (radiographic OA, two RCT's), **222 (92)**, 56	Anatomical phases	DIP, PIP, MCP	Summed total
		Anatomical lesions	DIP, PIP, MCP	Summed total
Verbruggen, 2012[25]	Secondary care (RCT), **60 (85)**, 61	Anatomical phases	DIP, PIP	No. of joints in each phase per patient
		GUSS	DIP, PIP	Summed total
Van 't Klooster, 2008[60]	Familial polyarticular OA (GARP), **40 (33)**, 60	OARSI	DIP, PIP, MCP	Score per joint
		Automatic JSW quantification	DIP, PIP, MCP	Mean score per joint
Vlychou, 2009[58]	Secondary care (OA patients), **22 (91)**, 63	Osteophytes, erosion (NVM)	DIP, PIP, IP1, MCP, CMC1	Present/absent per joint per feature
Wittoek, 2011[55]	Secondary care, erosive OA: **9 (67)**, median 61; non-erosive OA: **5 (100)**, median 63	Osteophytes, erosions (NVM)	DIP, PIP	Present/absent per joint per feature
Zhang, 2002[52]	General population (Framingham hand OA study), **1,032(64)**, age≥71	Modified KL	DIP, PIP, IP1, MCP, CMC1	Score per joint, present/absent of score ≥2 per patient

Abbreviations: AGES = Age, Gene/Environment Susceptibility, BLSA = Baltimore Longitudinal Study of AgingGARP = Genetics osteoarthritis and Progression, KLoSHA = Korean Longitudinal Study on Health and Aging, NVM = non-validated method, OA = osteoarthritis, .

Table III
Studies providing data on reliability of scoring methods ($n = 10$)

First author	No. of readers, *centers*	Intrareader reliability*	Interreader reliability*
Cross-sectional studies			
Addimanda[20]	2 (consensus), *1*	KL: ICC 0.994 Kallman: ICC 0.987, κ range per feature 0.42–0.81	N/A
Bijsterbosch[16]	3 (independent), *3*	KL: ICC range per reader 0.90–0.96 OARSI: ICC range per reader 0.77–0.97 Anatomical phases: ICC range per reader 0.88–0.97	KL: ICC range per two readers 0.84–0.91 OARSI: ICC range per two readers 0.80–0.95 Anatomical phases: ICC range per two readers 0.81–0.95
Haugen[21]	2 (independent), *2*	KL: ICC 0.97, κ 0.86 (one reader) OARSI (including marginal erosions): ICC range per feature 0.70–0.97, κ range per feature 0.75–0.88 (one reader)	KL: ICC 0.96, κ 0.79 OARSI (including marginal erosions): ICC range per feature 0.56–0.95, κ range per feature 0.62–0.81
Haugen[24]	2 (independent), *2*	KL: ICC 0.97, κ 0.82 (one reader) OARSI: ICC range per feature 0.62–0.96, κ range per feature 0.64–0.81 (one reader)	KL: ICC 0.95, κ 0.70 OARSI: ICC range per feature −0.07–0.94, κ range per feature 0.00.-0.77
Kallman[12]	4 independent, *2*	KL mean score: ICC 0.80, range per joint group 0.68–0.87 Kallman mean score: ICC per feature range 0.74–0.84, per feature per joint group range 0.62–0.93	KL mean score: ICC 0.74, range per joint group 0.74–0.81 Kallman mean score: ICC per feature range 0.29–0.71, per feature per joint group range 0.33–0.82
Kwok[22]	2 (consensus), *1*	OARSI (JSN): ICC 0.92 Semi-automated JSW: ICC 0.99, mean difference 0.017 mm (standard deviation (SD) 0.04), smallest detectable difference (SDD) 0.055 mm	N/A
Maheu[17]	2 (independent), *2*	KL: ICC range per reader 0.988–0.991 Kallman: ICC range per reader 0.962–0.999 Global: ICC range per reader 0.922–0.961 Anatomical phases: ICC range per reader 0.999–0.999	KL: ICC 0.951 Kallman: ICC 0.706 Global: ICC 0.859 Anatomical phases: ICC 0.996
Mancarella[23]	2, *not specified*	KL: ICC score per joint 0.99 Kallman: ICC score per joint 0.99	
Longitudinal studies			
Bijsterbosch[16]	3 (independent), *3* *Mean follow-up 2 years* *Mean follow-up 6 years*	KL: SDC range per reader 2.1–7.1 OARSI: SDC range per reader 1.2–10.2 Anatomical phases: SDC range per reader 1.4–7.8 KL: SDC range per reader 3.7–8.1 OARSI: SDC range per reader 3.0–11.1 Anatomical phases: SDC range per reader 3.5–9.9	KL: SDC 2.9 OARSI: SDC 4.1 Anatomical phases: SDC 2.7 KL: SDC 3.8 OARSI: SDC 4.6 Anatomical phases: SDC 4.0

(*continued on next page*)

Table III (*continued*)

First author	No. of readers, *centers*	Intrareader reliability*	Interreader reliability*
Haugen[24]	2 (independent), *2* *Mean follow-up 7 years*	KL: ICC 0.93, κ 0.83 (one reader) OARSI: ICC range per feature −0.02–0.96, κ range per feature 0.00–0.90 (one reader)	KL: ICC 0.83, κ 0.53 OARSI: ICC range per feature −0.03–0.90, κ range per feature −0.03–0.71
Kallman[12]	4 (independent), *2* *Mean follow-up 23 years*	N/A	KL: scattered agreement Deformity/collapse: agreement Cysts: disagreement Osteophytes/JSN/sclerosis: scattered agreement
Maheu[17]	2 (independent), *2* *Mean follow-up 1 year*	KL: ICC range per reader 0.990–0.998 Kallman: ICC range per reader 0.986–0.959 Global: ICC range per reader 0.939–0.956 Anatomical phases: ICC range per reader 0.941–0.988	KL: ICC 0.998 Kallman: ICC 0.995 Global: ICC 0.999 Anatomical phases: ICC 0.998
Verbruggen[26]	2 (independent), *1* *Mean follow-up 3 years*	Anatomical phases: agreement for two RCTs 84–93%, κ 0.6–0.8 Anatomical lesions: correlation for two RCTs *r* 0.7–0.9, R^2 44–87%	Anatomical phases: agreement for two RCTs 81–85%, κ 0.6–0.7 Anatomical lesions: correlation for two RCTs *r* 0.7–0.8, R^2 55–66%
Verbruggen[25]	2 (independent), *1* *Mean follow-up 1 year*	Anatomical phases: 96% agreement, κ 0.95 GUSS: ICC 0.97	Anatomical phases: 94% agreement, κ 0.92 GUSS: ICC 0.86, SDC 18

Abbreviations: κ = kappa, , N/A = not applicable, R^2 = explained variance.
* Unless stated otherwise ICCs are for summed total scores on patient level, κ's on joint level.

Table IV
Studies providing data on sensitivity to change of radiographic scoring methods in hand osteoarthritis ($n = 11$)

First author	Mean follow-up (years)	Definition of progression	Sequence known/ unknown	Results relevant for evaluation of sensitivity to change
Short-term				
Bijsterbosch[16]	2	Change > SDC	Known	*Percentage progression (range for three readers):* - KL: 19–56% - OARSI: 7–38% - Anatomical phases: 13–52%
Botha-Scheepers[27]	2	≥1 score	Known/ unknown	*Progression of JSN/osteophytes:* - chronological reading: 1/15% (SRM 0.38/0.41) - paired reading: 5/15% (SRM 0.00/0.39)
Botha-Scheepers[28]	2	≥1 score	Unknown	JSN: 19% progression, mean change 0.3, SRM 0.34 Osteophytes: 22% progression, mean change 0.4, SRM 0.35
Botha-Scheepers[29]	2	≥1 score	Unknown	JSN: 24% progression (≥2/≥3/≥4 score: 10/4/3%) Osteophytes: 22% progression (≥2/≥3/≥4 score: 10/4/3%)
Buckland-Wright[30]	1.5	Change > variations in precision	Not specified	JSW: 62% narrowing ($P < 0.02$) Subchondral sclerosis: 60% increase, 34% decrease Osteophytes: increase in size and no. ($P < 0.005$) Juxta-articular radiolucencies: increase in size ($P < 0.002$), not in no.
Maheu[17]	1	Change in summed score	Unknown	*SRM for two readers:* - KL: 0.17/0.24 - Kallman: 0.26/0.29 - Global: 0.17/0.27 - Anatomical phases: 0.18/0.27
Olejárová[31]	2	Change in summed score	Unknown	Erosive OA: change 5.0, $P > 0.05$ Non-erosive OA: change 4.3, $P > 0.05$
Verbruggen[26]	3	Change in anatomical phases, Change in anatomical lesions	Known	Anatomical lesions showed different progression between trial arms, anatomical phases did not.
Verbruggen[25]	1	Change in anato-mical N/S/J phase to E phase, Change in summed score	Unknown	*No. (%) joints with progression to E phase:* - Total group: 24 (2.8%) of 848 N/S/J joints - Placebo treated: 15 (3.6%) of 429 N/S/J joints - Adalimumab treated: 9 (2.1%) of 419 N/S/J joints Mean difference GUSS (baseline palpable swelling yes/no): - Placebo: −5/3 - Adalimumab: 4/1
Long-term				
Bijsterbosch[16]	6	Change > SDC	Known	*Percentage progression (range for three readers):* - KL: 51–80% - OARSI: 33–74% - Anatomical phases: 27–66%
Haugen[24]	7.3	Change in score	Known	*Progression (percentage of joints):* - KL: 29% - OARSI: osteophytes 19%, JSN 13%, erosions 9%, mala-lignment 4%, cysts 2%, sclerosis 1%
Verbruggen[14]	4.6	Change in anatomical phases, Change in anatomical lesions	Known	Progression of anatomical lesions more frequent in PIP/DIP than MCP. Progression of anatomical phases in 43%. Progression according anatomical phases and anatomical lesions yielded comparable results.

Table V
Studies providing data on feasibility of radiographic scoring methods in hand osteoarthritis ($n = 4$)

First author	No. of radiographs	Mean (SD) time to perform scoring
Bijsterbosch[16]	3	KL: 4.3 (2.5) min OARSI: 9.3 (6.0) min Anatomical phases: 2.8 (1.5) min
Kessler[11]	1	Kessler: 5 min per hand Kallman: 10–15 min per hand Lane: 10–15 min per hand
Kwok[22]	1	Semi-automated JSW measurement: 5.1 (2.8) min
Maheu[17]	1	KL: 1.9 (0.6) min Kallman: 3.5 (0.7) min Global score: 1.5 (0.5) min Anatomical phases: 1.6 (0.5) min

Abbreviations: min = minutes, no. = number.

respectively)[20]. Rees *et al.* examined the association between KL and OARSI scores and clinical nodes, reporting ORs only for the KL method (range per joint 2.3–21.2). Regarding the OARSI atlas, JSN was mentioned to be more strongly associated with clinical nodes than osteophytes.[41]

Seventeen studies assessed clinical symptoms and hand function in comparison to radiographic scoring methods (KL: $n = 14$, OARSI: $n = 3$, Kallman: $n = 1$, JWS/JSN: $n = 1$)[22,24,33,36,37,39,43–52,62]. All studies reported significant associations between radiographic OA features and pain and disability, of which four showed a dose-dependent association between KL and OARSI scores and pain[24,43,44,48]. Of the nine studies assessing grip or pinch strength, only two did not find an association with radiographic OA (1x KL, 1x JSW/JSN, analyzed on patient level).[22,50]

Only one study assessed longitudinal data, showing incident or progressive KL or OARSI scores to be associated with incident pain

Table VI
Studies providing data on validity of scoring methods ($n = 37$)

First author	Validation method	Results relevant for evaluation of validity
Clinical: structural findings at physical examination		
Addimanda[20]	Heberden/Bouchard nodes (yes/no)	*OR (95% CI) for nodes on joint level, adjusted for disease duration, body mass index (BMI):* - KL: 2.20 (2.09, 2.31) - Kallman: 1.17 (1.62, 1.72) - Kallman JSN: 2.57 (2.40, 2.75) - Kallman osteophytes: 3.19 (2.97, 3.42) - Kallman central erosions: 7.4 (6.0, 10.1)
Bagge[33]	Nodes/periarticular enlarge-ment, instability, squaring (yes/no ≥1 feature per joint)	Correlated with KL score in all joint groups (correlation coefficient not provided), test for linear trend: $P < 0.01$. Clinical features also present in KL 0 joint groups.
Caspi[34]	Nodes, malalignment DIP/PIP (summed)	*Correlation with OARSI:* - summed total: r 0.4 (P 0.001) - DIP/PIP: range per joint r 0.18–0.52 (P 0.004–0.0001)
Cicuttini[35]	Heberden nodes (yes/no)	κ with DIP osteophytes (Burnett): 0.36 (95% CI 0.33, 0.39)
Hart[36]	Nodes (yes/no)	Sensitivity for KL ≥2: range per joint group 19–49% Specificity for KL ≥2: range per joint group 87–98%
Hart[37]	Nodes IP (graded 0–4), squaring CMC1 (grade 0–1)	Prevalence node ≥2: KL0: 3%, KL1: 19%, KL2: 48%, KL3: 74%, KL4: 82% Prevalence squaring: KL0: 5%, KL1: 11%, KL2: 25%, KL3: 41%, KL4: 70% (correlation coefficient not specified)
Jonsson[38]	Nodes, deformity (graded 0–3, summed)	Correlation summed score with summed total KL: males r 0.47, females r 0.66 Prevalence KL ≥ 2 (DIP 67%, PIP 32%, CMC1 20%) higher as compared to clinical grade ≥2 (DIP 54%, PIP 19%, CMC1 10%)
Kwok[22]	Nodes (yes/no)	*β (95% CI) for nodes on joint level, adjusted for age, sex, BMI, family effect, mean phalanx width:* - JSW: −0.37 (−0.40, −0.34) - JSN: 0.48 (0.42, 0.55)
Marshall[39]	Nodes, deformity, enlargement (yes/no)	*OR (95% CI) of presence of ≥1 feature for:* - KL ≥ 2 in CMC1: 2.2 (1.5, 3.3) - KL ≥ 2 in any thumb joint: 3.1 (2.1, 4.5)
Mathiessen[40]	Nodes (yes/no)	Osteophytes (OARSI) in 30% of joints, nodes in 37% of joints
Rees[41]	Nodes (yes/no)	KL ≥ 2 associated with any node on patient level: OR range per joint 2.26–21.23 (adjusted for age, sex, BMI, hand dominance, trauma, occupation, sports) JSN/osteophytes (OARSI) also associated with nodes ($P < 0.001$); ORs of JSN greater than ORs of osteophytes in all joints except for IP1/CMC1
Stern[42]	Nodes (yes/no)	Sensitivity for KL ≥ 2: range per joint group 42–100% Specificity for KL ≥ 2: range per joint group 17–94%
Clinical: symptoms, function		
Bagge[33]	Pain/stiffness (interview, yes/no)	Correlated with KL score in all joint groups (correlation coefficient not provided), test for linear trend: $P < 0.01$.
Ceceli[62]	Pain (visual analog scale(VAS)), disability (Disabilities of the Arm Shoulder and Hand (DASH) questionnaire), dexterity (Purdue pegboard test), grip/pinch strength	*Correlation with summed Kallman score right/left hand:* - Pain: r 0.17/0.18 ($P > 0.05$) - Disability: r 0.29/0.30 ($P < 0.05$) - Dexterity: r −0.26/−0.30 ($P < 0.05$) - Grip strength: r −0.37/−0.40 ($P < 0.05$) - Pinch strength: r range per test −0.31 to −0.25/−0.35 to −0.27 ($P < 0.05$)
Dahaghin[43]	Pain (interview, yes/no)/disability (Health Assessment Questionnaire (HAQ))	*OR (95% CI) for KL ≥ 2/≥3/4 on patient level, adjusted for age, sex:* - pain: 1.9 (1.5, 2.4)/1.8 (1.3, 2.5)/3.6 (2.2, 5.8) - disability: 1.5 (1.1, 2.1)/1.6 (1.1, 2.5)/1.6 (0.9, 2.9) Pain associated with KL ≥ 2 in PIP/CMC1/STT, disability with KL ≥ 2 in MCP Adjusted OR (95% CI) for KL ≥ 2 in all joint groups: pain 2.7 (1.4, 5.2), disability 2.7 (1.3, 6.0)

(*continued on next page*)

Table VI (*continued*)

First author	Validation method	Results relevant for evaluation of validity
Ding[44]	Pain (questionnaire, yes/no per joint, summed)	Correlation with summed total KL: r 0.26 (P 0.0005) Correlation with no. KL $\geq$ 2 joints: r 0.28 (P 0.0005) *prevalence ratio (PR) (95% CI) for pain on joint level, adjusted for age, occupation:* - KL 2: 1.70 (1.44, 2.01) - KL $\geq$ 3: 5.17 (4.34, 6.16) *Adjusted PR (95% CI) for mild/moderate pain on joint level:* - KL 2: 1.93 (1.54, 2.41)/2.21 (1.58, 3.10) - KL $\geq$ 3: 4.92 (3.77, 6.43)/11.73 (8.95, 15.38)
Dominick[45]	Grip/pinch strength	*β (P-value) for grip/pinch strength, adjusted for age, sex, pain, chondro-calcinosis, hand hypermobility:* - Summed total KL: −0.67 (<0.001)/−0.16 (<0.001) - KL $\geq$ 2 PIP: −6.67 (0.003)/−1.17 (0.070) - KL $\geq$ 2 MCP: −3.32 (0.114)/−1.78 (0.003) - KL $\geq$ 2 CMC: −9.06 (<0.001)/−1.03 (0.049) - KL $\geq$ 2 per finger: range −1.81 to −11.08 ($P < 0.05$)
El-Sherif[46]	AUSCAN, morning stiffness (minutes), grip strength, Ritchie index	AUSCAN pain/function higher in KL4 than KL2 ($P < 0.05$) *Correlation with KL score:* - AUSCAN pain: r 0.459 (P 0.003), function: r 0.394 (P 0.012) - Grip strength right hand: r −0.322 (P 0.043) Other measures not significantly correlated with KL
Hart[36]	Tenderness, pain on movement (physical examination, yes/no)	*Comparison tenderness/pain on movement with KL $\geq$ 2:* - sensitivity: range per joint group 7–26%/1–22% - specificity: range per joint group 92–99%/96–99%
Hart[37]	Pain, stiffness (interview, yes/no)	Prevalence symptoms in patients with KL < 2: 15%, KL2: 49%, KL3-4: 81%; test for linear trend: $P < 0.01$
Haugen[24]	Tenderness on palpation (yes/no), grip strength, AUSCAN	*Cross-sectional OR (95% CI) for tenderness on joint level, adjusted for age, sex:* - KL score 1/2/3/4: 1.4 (1.2, 1.7)/3.0 (2.4, 3.7)/6.8 (4.5, 10)/5.3 (3.3, 8.6) - OARSI osteophytes score 1/2/3: 2.8 (2.3, 3.4)/4.3 (3.0, 6.3)/4.5 (2.9, 7.0) - OARSI JSN score 1/2/3: 0.9 (0.7, 1.2)/1.9 (1.4, 2.5)/2.5 (1.7, 3.7) - OARSI erosions: 3.3 (2.3, 4.9), malalignment: 2.8 (2.0, 3.9), cysts: 2.2 (1.4,3.3), sclerosis: 2.6 (1.1, 6.0) AUSCAN pain associated with summed KL and OARSI osteophytes/JSN. AUSCAN function associated with summed KL and OARSI osteophytes, JSN, erosions, cysts. Grip strength associated with summed KL and all OARSI features except for sclerosis. Summed KL per joint group only associated with grip strength (CMC1 strongest) *Adjusted OR (95% CI) of progressive/incident scores for incident tenderness:* - KL score 1/2/3/4: 1.2 (0.7, 2.0)/1.5 (0.9, 2.4)/5.7 (3.0, 11)/11 (4.0, 33) - OARSI osteophytes: 3.0 (2.0, 4.4), JSN: 2.8 (1.7, 4.7), erosions: 8.4 (4.7, 15), malalignment: 3.8 (1.9, 7.4), cysts: 2.2 (0.9, 5.0), sclerosis: 2.4 (0.8, 8.0) Increasing summed KL and OARSI JSN/malalignment associated with increased AUSCAN function. More malalignment associated with less grip strength Change summed KL per joint group not associated with AUSCAN/grip strength
Jones[47]	AUSCAN, grip strength	*Association with summed OARSI per joint group, adjusted for age/sex/other joints/ Heberden nodes:* - AUSCAN pain: PIP β 0.17, CMC1 β 0.14 ($P < 0.05$) - AUSCAN function: PIP β 0.15, CMC1 β 0.19 ($P < 0.05$) - grip strength: PIP β −0.12, CMC1 β −0.09 ($P < 0.05$)
Kortekaas[48]	AUSCAN, pain (VAS), Doyle index of hands	*OR (95% CI) for pain on palpation on joint level, adjusted for age, sex, BMI:* - osteophytes score 1/2/3: 2.2 (1.7, 2.9)/3.9 (2.6, 5.9)/4.8 (2.7, 8.4) - JSN score 1/2/3: 2.0 (1.4, 2.8)/5.3 (3.1, 9.1)/6.4 (2.7, 14.8) Summed osteophytes/JSN not associated with AUSCAN pain, VAS, Doyle.
Kwok[22]	AUSCAN, pain on palpation (yes/no), grip strength, mobility	*β (95% CI) for JSW/JSN on joint level, adjusted for age, sex, BMI, family effect, mean phalanx width:* - self-reported pain: −0.21 (−0.27, −0.16)/0.39 (0.30, 0.48) - pain on palpation: −0.25 (−0.29, −0.21)/0.37 (0.29, 0.44) No. joints with self-reported pain/pain on palpation, AUSCAN pain/function and mobility associated with summed JSW/JSN. Grip strength not associated
Lee[49]	Grip/pinch strength, disability (DASH questionnaire)	*Associations with summed KL, adjusted for age/sex ($P < 0.05$):* - grip strength: thumb β −1.05, third finger β −2.17 - pinch strength: thumb β −0.28, second finger β −0.26 -disability: thumb β 1.53, second finger β 0.63, third finger β 3.97
Marshall[39]	AUSCAN, pain during activity/pain in past month (questionnaire, yes/no), grip/pinch strength, grind test, Finkelstein's test	*OR (95% CI) for KL $\geq$ 2 in CMC1/any thumb joint:* - Pain during activity: 2.1 (1.5, 2.9)/2.2 (1.6, 3.2) - Pain in past month: 1.5 (1.0, 2.1)/1.4 (1.0, 2.0) - Grind test: 1.8 (1.1, 2.9)/1.7 (1.0, 2.9), Finkelstein's test not associated
Ozkan[50]	Grip/pinch strength, Dreiser's functional index, disability (HAQ)	Disability KL score <2/2/3-4: 2.40/2.10/6.45 (KL3-4 vs KL < 2/2 $P < 0.05$) Dreiser's index KL score <2/2/3-4: 2.73/2.10/9.25 (KL3-4 vs KL < 2/2 $P < 0.05$) Grip/pinch strength not different between KL scores

Table VI (*continued*)

First author	Validation method	Results relevant for evaluation of validity
Sonne-Holm[51]	Pain CMC1 (interview, yes/no)	*OR (95% CI) for pain, adjusted for age, sex, BMI:* - KL: 1.48 (1.33, 1.65) - Sclerosis/cyst: 1.48 (1.23, 1.77)/1.23 (1.03, 1.47)
Zhang[52]	Functional limitations (questionnaire), grip strength	JSW and osteophytes not associated. Patients with KL $\geq$ 2 and joint pain/aching/stiffness had more functional limitations and lower grip strength; age adjusted difference (95% CI) men 3.1 kg (1.8, 4.4), women 1.9 kg (1.4, 2.4)
Histological		
Sunk[53,69]	Modified Mankin score (range 0–14; >5 = OA)	Correlation with KL score (DIP/PIP): r 0.87/0.79 ($P < 0.0001$) Correlation with OARSI JSN: r 0.77/0.76, osteophytes: r 0.89/0.69 ($P < 0.0001$) Sensitivity KL $\geq$ 2 for Mankin >5 (DIP/PIP): 84.6/54.2%, specificity: 100/100%
MRI		
Drape[32]	Pedicled cysts DIP (yes/no)	19 pedicled cysts: 16 associated with osteophytes/JSN on CR, three no osteophytes/JSN on CR
Grainger[54]	Erosions (central/marginal, yes/no)	37 MRI erosions: 24% also on CR (44% of central, 5% of marginal erosions) All CR erosions also on MRI
Haugen[21]	Oslo hand OA score (graded per feature)	Agreement with osteophytes κ 0.41, JSN κ 0.50, central erosions κ 0.75, central/marginal erosions κ 0.43, cysts κ 0.11, malalignment κ 0.50
Wittoek[55]	Erosions, osteophytes (yes/no)	Prevalence erosions: MRI PIP 29%, DIP 68%, CR PIP 11%, DIP 38% PIP osteophytes (erosive/non-erosive) hand OA MRI 25/50%, CR 42/40% DIP osteophytes: MRI and CR >80%
CT		
Saltzherr[61]	JSN, osteophytes, subchon-dral sclerosis, cyst, erosion, subluxation (OA defined on no. of features)	Prevalence of individual features and OA higher according to CT than CR
US		
Iagnocco[56]	Erosions (yes/no)	US erosions in 16 (72.7%) of 22 CR erosive hand OA patients. No US erosions in CR classical hand OA patients ($n = 88$).
Keen[57]	JSN, osteophytes (yes/no)	Osteophytes: κ 0.54 (77.8% agreement) JSN: κ 0.436 (74.6% agreement)
Kortekaas[48]	Osteophytes (yes/no)	US osteophytes 69%, OARSI osteophytes 46%
Mancarella[23]	Cartilage thickness (mm)	Negatively correlated with KL and Kallman score ($P < 0.0001$)
Mathiessen[40]	Osteophytes (yes/no)	OARSI osteophytes in 30% of joints, US osteophytes in 53% of joints CR and US: 57.3% exact agreement, 88.3% close agreement
Vlychou[58]	Central erosions, osteophytes (yes/no)	CR detected less erosions/osteophytes (17/47%) than US (35/55%), $P < 0.05$ Difference most apparent in DIP and PIP
Wittoek[55]	Erosions, osteophytes (yes/no)	CR detected less erosions (PIP 11%, DIP 38%) than US (21, 52%) in erosive and non-erosive hand OA CR detected less PIP osteophytes (41%) than US (54%). CR and US both detected >80% DIP osteophytes
Digital photography		
Jones[47]	Heberden nodes (yes/no)	Correlation with OARSI score $\geq$1 in DIP joints: r 0.74 ($P < 0.001$)
Jonsson[38]	Tissue enlargement/deformity (graded 0–3 per joint, summed)	Prevalence OA higher according to KL $\geq$ 2 (DIP 67%, PIP 32%, CMC1 20%) as compared to digital photograph $\geq$2 (DIP 33%, PIP 20%, CMC1 3%) Correlation summed score with summed total KL: males r 0.35, females r 0.53
Stern[42]	Hard tissue enlargement (yes/no)	Sensitivity for KL $\geq$ 2: range per joint 17–74% Specificity for KL $\geq$ 2: range per joint 67–92%
Other measures of JSW		
Huetink[59]	True JSW by micrometer	Compared to automatic JSN quantification: Mean difference (SD): phantom joints: 0.052 (0.014) mm, cadaver joints: 0.210 (0.115) mm SDD: phantom joints 0.028 mm, cadaver joints: 0.226 mm
van't Klooster[60]	Automatic JSW quantification (mm)	Association with OARSI JSN: R^2 0.54, $P < 0.01$

Abbreviations: kg = kilogram, r = correlation coefficient.

on joint level and with change in Australian/Canadian Hand Osteoarthritis Index (AUSCAN) pain/function and grip strength.[24]

One study examined the association between the KL and OARSI scoring methods and histological findings on joint group level, showing a good correlation ($r \geq 0.7$) as well as a high sensitivity and specificity.[53]

Four studies assessed individual features of hand OA by both radiography and MRI[21,32,54,55]. The agreement between the two methods was lowest for the presence of cysts and highest for central erosions[21]. Three of the studies showed that MRI detected more osteophytes, cysts and erosions as compared to radiography.[32,54,55]

One study assessed individual features of CMC1 and STT OA by both radiography and CT, reporting the latter to detect more JSN, osteophytes, subchondral sclerosis, cysts, erosions and subluxation.[61]

Seven studies used both US and radiography to assess hand OA signs[23,40,48,55–58]. Six of the studies examined individual radiographic features and reported US to detect more osteophytes and erosions than radiography. A study on KL and Kallman scores reported a negative correlation between radiographic JSN and US-detected cartilage thickness on joint level.[23]

Three studies examined hand OA using digital photography and radiography[38,42,47]. Two studies, performed on joint group level, reported a good correlation between OARSI scores and Heberden nodes on digital photography ($r = 0.74$), and a weak to moderate correlation between summed KL scores and summed digital photograph score (comprising enlargement and deformity) on digital photography (males $r = 0.35$, females $r = 0.53$).[38,47]

Finally, two studies examined quantitative measures of JSW, both on individual joint level[59,60]. Van't Klooster *et al.* showed that automatic JSW quantification was associated with JSN scored

according to the OARSI atlas[60]. Huetink *et al.* reported that automatic JSW quantification has a high accuracy in measuring the true JSW (assessed by micrometer).[59]

Discussion

This review aimed at evaluating the radiographic scoring methods used in hand OA research and to assess their metric properties. We noticed that a wide variety of scoring methods has been used in studies evaluating radiographic hand OA. Furthermore, the joints that were examined and the analysis of the obtained scores differed extensively across studies. Evaluation of metric properties of the evaluated scoring methods regarding reliability, sensitivity to change, feasibility and validity did not reveal major differences.

Both intra- and interreader reliability of all evaluated radiographic scoring methods were good for summed scores and global scores, for both cross-sectional and longitudinal radiographic scoring. When grading individual radiographic features, the highest reliability was reported for the scoring of erosions and osteophytes and the lowest for the scoring of cysts.

When evaluating sensitivity to change, only one study evaluated this in different groups of patients (trial arms) using different scoring methods. Although such comparative studies may provide the best insights in sensitivity to change, the included observational follow-up studies showed the ability to detect change in structural damage over time with CR. Change over time was observed even in short term follow-up studies (<3 years). Reported SRMs were similar for all evaluated scoring methods.

The feasibility of scoring methods has been described in a limited number of studies. The performance time of the scoring differed not only across the evaluated scoring method but also across studies, and was shown to increase with the amount of structural damage.

A large number of studies investigated the validity of radiographic OA findings in comparison with clinical findings at physical examination (such as nodes and deformities) and symptoms and function; there was large variation between these studies. This could be due to the various analyses of radiographic and clinical findings, e.g., joint level vs patient level, and individual features vs summed scores. Furthermore, studies were difficult to compare because of the use of different effect measures, such as odds ratios (ORs), correlation coefficients, sensitivity and specificity. In general we can say that there was moderate agreement between radiographic features and structural findings at physical examination. The association of radiographic findings with hand function and symptoms was reported to be stronger than the association with findings at physical examination. All evaluated radiographic scores were associated with grip strength and pain, the relation with pain was observed on joint level as well as on patient level, and was shown to be dependent on the radiographic severity. No differences between the evaluated radiographic scoring methods were observed. Only few studies assessed longitudinal associations between radiography and pain or function, requiring further validation.

In comparison with other imaging methods, radiography appeared to detect fewer structural damage than MRI, CT and US, and more structural damage than digital photography. However, the findings on MRI, CT and digital photography require further confirmation because of limited evidence. Agreement between radiography and other imaging methods was assessed most often on joint level and differed per feature.

Although no major differences regarding the metric properties of the evaluated radiographic scoring methods were observed in this review, the examined joints and analysis of the obtained scores were shown to differ extensively across studies. All kinds of presentation of radiographic outcome measures were used, such as scores per joint, summed scores, presence/absence of radiographic OA features, or the highest scored joint. Summed scores were used most frequently for evaluation of the reliability of radiographic scoring methods and change of structural damage over time, analyzed on patient level. When evaluating the validity of scoring methods, analyses on individual joint level or on joint group level were performed most often.

The various examined joints within hand OA research has been described before in a review by Marshall *et al.* In addition, they evaluated the use of definitions of hand OA, reporting some agreement in the definition of individual joint OA but a wide variation in defining overall hand OA[65]. Kerkhof *et al.* showed that the use of varying definitions of radiographic OA within the same study leads to different results[66]. Therefore, as stated before by Haugen *et al.*, standardization of the evaluation and definition of radiographic hand OA with respect to scoring methods, examined joints and required number of affected joints could reduce the variation across studies.[67]

Based on this review, it is not possible to decide on what radiographic scoring method should be recommended in hand OA research. Although no major differences regarding metric properties of the scoring methods were observed, the amount of supporting evidence differed for the evaluated methods, which may provide an argument for recommendation of specific scoring methods. Most evidence across all evaluated domains is available for the KL and OARSI scoring methods. Although global scoring methods may be more reliable than the scoring of individual radiographic features, individual features may be more suitable for evaluation of specific study objectives. Therefore, the OARSI scoring method may be recommended for evaluation of individual radiographic features in addition to use of the KL scoring method for global radiographic assessment. The OARSI Task Force recommendations for the design and conduct of clinical trials in hand OA already stated that the use of either aggregate radiographic scores or grading of individual features depends on the aim of study[9]. However, consensus should be reached on a more specific definition; when should a global or individual feature score be used and what specific scoring method should be recommended. Furthermore, consensus on the evaluated joints, presentation of the radiographic outcome measures and the definition of hand OA will help to enhance the comparability of studies in hand OA.

A limitation of this study is that the methodological quality of the included studies was not assessed, due to the heterogeneity across studies regarding their purpose. The heterogeneity regarding examined joints and analyses of obtained radiographic scores did not enable performance of a meta-analysis. Furthermore, publication bias was not addressed.

Although we aimed to provide a comprehensive overview of available literature, the formulated inclusion and exclusion criteria resulted in a specific selection of studies.

Consequently, some radiographic scoring methods were not included in this review, being the Eaton-Littler classification system and the recently developed interphalangeal OA radiographic simplified (iOARS) score. These methods have not been evaluated for reliability together with another method.[68,69]

Since sensitivity to change was evaluated in follow-up studies assessing hand OA by at least two radiographic scoring methods in case of long-term follow-up studies (>3 year), a number of studies or abstracts evaluating change in KL and OARSI scores could not be included.[3,70–72]

In the evaluation of the feasibility of the available radiographic scoring methods in hand OA, we did not focus on the importance of

radiographic techniques. Dela Rosa *et al.* evaluated the reliability of scoring OA of the CMC1s according to the Eaton method when using different X-ray views, showing that a combination of the posterior-anterior, lateral and Bett's view showed a higher reliability than using only one or two views[73]. Standardization of radiographic techniques might further enhance comparability of studies in hand OA.

In conclusion, this systematic review provides an overview of the radiographic scoring methods used in the assessment of structural damage in hand OA. We showed that several scoring methods are available, evaluation of their metric properties regarding reliability, sensitivity to change, feasibility and validity did not reveal major differences. The examined joints and analysis of the obtained radiographic scores differed extensively across all studies. To enhance comparability across studies in hand OA, consensus has to be reached on a preferred scoring method, as well as on the examined joints and the used outcome measure.

Contributions

Authors made substantial contributions to the following: (1a) conception and design of the study: AWV, PB, DMH, MK; (1b) acquisition of data: AWV, JWS, MK; (1c) analysis and interpretation of data: AWV, PB, IKH, DMH, FRR, MK (2) drafting or critically revising of manuscript: AWV, PB, IKH, JWS, DMH, FRR, MK; (3) final approval of manuscript: AWV, PB, IKH, JWS, DMH, FRR, MK.

Competing interest statement

There were no competing interests.

Funding

This work was supported by the Dutch Arthritis Foundation (grant number 10-1-309).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.joca.2014.05.026.

References

1. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58:26–35.
2. van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989;48:271–80.
3. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, *et al.* Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Ann Rheum Dis 2011;70:1581–6.
4. Mahendira D, Towheed TE. Systematic review of non-surgical therapies for osteoarthritis of the hand: an update. Osteoarthritis Cartilage 2009;17:1263–8.
5. Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, *et al.* EULAR evidence based recommendations for the management of hand osteoarthritis: report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66:377–88.
6. Towheed TE. Systematic review of therapies for osteoarthritis of the hand. Osteoarthritis Cartilage 2005;13:455–62.
7. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
8. Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, *et al.* Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. Osteoarthritis Cartilage 1996;4:217–43.
9. Maheu E, Altman RD, Bloch DA, Doherty M, Hochberg M, Mannoni A, *et al.* Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. Osteoarthritis Cartilage 2006;14:303–22.
10. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
11. Kessler S, Dieppe P, Fuchs J, Sturmer T, Gunther KP. Assessing the prevalence of hand osteoarthritis in epidemiological studies. The reliability of a radiological hand scale. Ann Rheum Dis 2000;59:289–92.
12. Kallman DA, Wigley FM, Scott Jr WW, Hochberg MC, Tobin JD. New radiographic grading scales for osteoarthritis of the hand. Reliability for determining prevalence and progression. Arthritis Rheum 1989;32:1584–91.
13. Altman RD, Hochberg M, Murphy Jr WA, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage 1995;3(Suppl A):3–70.
14. Verbruggen G, Veys EM. Numerical scoring systems for the anatomic evolution of osteoarthritis of the finger joints. Arthritis Rheum 1996;39:308–20.
15. Verbruggen G, Wittoek R, Vander CB, Elewaut D. Morbid anatomy of 'erosive osteoarthritis' of the interphalangeal finger joints: an optimised scoring system to monitor disease progression in affected joints. Ann Rheum Dis 2010;69:862–7.
16. Bijsterbosch J, Haugen IK, Malines C, Maheu E, Rosendaal FR, Watt I, *et al.* Reliability, sensitivity to change and feasibility of three radiographic scoring methods for hand osteoarthritis. Ann Rheum Dis 2011;70:1465–7.
17. Maheu E, Cadet C, Gueneugues S, Ravaud P, Dougados M. Reproducibility and sensitivity to change of four scoring methods for the radiological assessment of osteoarthritis of the hand. Ann Rheum Dis 2007;66:464–9.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
19. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in Rheumatology. J Rheumatol 1998;25:198–9.
20. Addimanda O, Mancarella L, Dolzani P, Punzi L, Fioravanti A, Pignotti E, *et al.* Clinical and radiographic distribution of structural damage in erosive and nonerosive hand osteoarthritis. Arthritis Care Res (Hoboken) 2012;64:1046–53.
21. Haugen IK, Boyesen P, Slatkowsky-Christensen B, Sesseng S, Bijsterbosch J, van der Heijde D, *et al.* Comparison of features by MRI and radiographs of the interphalangeal finger joints in patients with hand osteoarthritis. Ann Rheum Dis 2012;71:345–50.
22. Kwok WY, Bijsterbosch J, Malm SH, Biermasz NR, Huetink K, Nelissen RG, *et al.* Validity of joint space width measurements in hand osteoarthritis. Osteoarthritis Cartilage 2011;19:1349–55.
23. Mancarella L, Magnani M, Addimanda O, Pignotti E, Galletti S, Meliconi R. Ultrasound-detected synovitis with power Doppler

signal is associated with severe radiographic damage and reduced cartilage thickness in hand osteoarthritis. Osteoarthritis Cartilage 2010;18:1263–8.
24. Haugen IK, Slatkowsky-Christensen B, Boyesen P, van der Heijde D, Kvien TK. Cross-sectional and longitudinal associations between radiographic features and measures of pain and physical function in hand osteoarthritis. Osteoarthritis Cartilage 2013;21:1191–8.
25. Verbruggen G, Wittoek R, Vander CB, Elewaut D. Tumour necrosis factor blockade for the treatment of erosive osteoarthritis of the interphalangeal finger joints: a double blind, randomised trial on structure modification. Ann Rheum Dis 2012;71:891–8.
26. Verbruggen G, Goemaere S, Veys EM. Systems to assess the progression of finger joint osteoarthritis and the effects of disease modifying osteoarthritis drugs. Clin Rheumatol 2002;21:231–43.
27. Botha-Scheepers S, Watt I, Breedveld FC, Kloppenburg M. Reading radiographs in pairs or in chronological order influences radiological progression in osteoarthritis. Rheumatology (Oxford) 2005;44:1452–5.
28. Botha-Scheepers S, Riyazi N, Watt I, Rosendaal FR, Slagboom E, Bellamy N, *et al*. Progression of hand osteoarthritis over 2 years: a clinical and radiological follow-up study. Ann Rheum Dis 2009;68:1260–4.
29. Botha-Scheepers SA, Watt I, Slagboom E, Meulenbelt I, Rosendaal FR, Breedveld FC, *et al*. Influence of familial factors on radiologic disease progression over two years in siblings with osteoarthritis at multiple sites: a prospective longitudinal cohort study. Arthritis Rheum 2007;57:626–32.
30. Buckland-Wright JC, Macfarlane DG, Lynch JA, Clark B. Quantitative microfocal radiographic assessment of progression in osteoarthritis of the hand. Arthritis Rheum 1990;33:57–65.
31. Olejarova M, Kupka K, Pavelka K, Gatterova J, Stolfa J. Comparison of clinical, laboratory, radiographic, and scintigraphic findings in erosive and nonerosive hand osteoarthritis. Results of a two-year study. Jt Bone Spine 2000;67:107–12.
32. Drape JL, Idy-Peretti I, Goettmann S, Salon A, Abimelec P, Guerin-Surville H, *et al*. MR imaging of digital mucoid cysts. Radiology 1996;200:531–6.
33. Bagge E, Bjelle A, Eden S, Svanborg A. Osteoarthritis in the elderly: clinical and radiological findings in 79 and 85 year olds. Ann Rheum Dis 1991;50:535–9.
34. Caspi D, Flusser G, Farber I, Ribak J, Leibovitz A, Habot B, *et al*. Clinical, radiologic, demographic, and occupational aspects of hand osteoarthritis in the elderly. Semin Arthritis Rheum 2001;30:321–31.
35. Cicuttini FM, Baker J, Hart DJ, Spector TD. Relation between Heberden's nodes and distal interphalangeal joint osteophytes and their role as markers of generalised disease. Ann Rheum Dis 1998;57:246–8.
36. Hart DJ, Spector TD, Brown P, Wilson P, Doyle DV, Silman AJ. Clinical signs of early osteoarthritis: reproducibility and relation to X ray changes in 541 women in the general population. Ann Rheum Dis 1991;50:467–70.
37. Hart D, Spector T, Egger P, Coggon D, Cooper C. Defining osteoarthritis of the hand for epidemiological studies: the Chingford Study. Ann Rheum Dis 1994;53:220–3.
38. Jonsson H, Helgadottir GP, Aspelund T, Sverrisdottir JE, Eiriksdottir G, Sigurdsson S, *et al*. The use of digital photographs for the diagnosis of hand osteoarthritis: the AGES-Reykjavik study. BMC Musculoskelet Disord 2012;13:20.
39. Marshall M, van der Windt D, Nicholls E, Myers H, Dziedzic K. Radiographic thumb osteoarthritis: frequency, patterns and associations with pain and clinical assessment findings in a community-dwelling population. Rheumatology (Oxford) 2011;50:735–9.
40. Mathiessen A, Haugen IK, Slatkowsky-Christensen B, Boyesen P, Kvien TK, Hammer HB. Ultrasonographic assessment of osteophytes in 127 patients with hand osteoarthritis: exploring reliability and associations with MRI, radiographs and clinical joint findings. Ann Rheum Dis 2013;72:51–6.
41. Rees F, Doherty S, Hui M, Maciewicz R, Muir K, Zhang W, *et al*. Distribution of finger nodes and their association with underlying radiographic features of osteoarthritis. Arthritis Care Res (Hoboken) 2012;64:533–8.
42. Stern AG, Moxley G, Sudha Rao TP, Disler D, McDowell C, Park M, *et al*. Utility of digital photographs of the hand for assessing the presence of hand osteoarthritis. Osteoarthritis Cartilage 2004;12:360–5.
43. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis 2005;64:682–7.
44. Ding H, Solovieva S, Vehmas T, Riihimaki H, Leino-Arjas P. Finger joint pain in relation to radiographic osteoarthritis and joint location–a study of middle-aged female dentists and teachers. Rheumatology (Oxford) 2007;46:1502–5.
45. Dominick KL, Jordan JM, Renner JB, Kraus VB. Relationship of radiographic and clinical variables to pinch and grip strength among individuals with osteoarthritis. Arthritis Rheum 2005;52:1424–30.
46. El-Sherif HE, Kamal R, Moawyah O. Hand osteoarthritis and bone mineral density in postmenopausal women; clinical relevance to hand function, pain and disability. Osteoarthritis Cartilage 2008;16:12–7.
47. Jones G, Cooley HM, Bellamy N. A cross-sectional study of the association between Heberden's nodes, radiographic osteoarthritis of the hands, grip strength, disability and pain. Osteoarthritis Cartilage 2001;9:606–11.
48. Kortekaas MC, Kwok WY, Reijnierse M, Huizinga TW, Kloppenburg M. Osteophytes and joint space narrowing are independently associated with pain in finger joints in hand osteoarthritis. Ann Rheum Dis 2011;70:1835–7.
49. Lee HJ, Paik NJ, Lim JY, Kim KW, Gong HS. The impact of digit-related radiographic osteoarthritis of the hand on grip-strength and upper extremity disability. Clin Orthop Relat Res 2012;470:2202–8.
50. Ozkan B, Keskin D, Bodur H, Barca N. The effect of radiological hand osteoarthritis on hand function. Clin Rheumatol 2007;26:1621–5.
51. Sonne-Holm S, Jacobsen S. Osteoarthritis of the first carpometacarpal joint: a study of radiology and clinical epidemiology. Results from the Copenhagen Osteoarthritis Study. Osteoarthritis Cartilage 2006;14:496–500.
52. Zhang Y, Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. Prevalence of symptomatic hand osteoarthritis and its impact on functional status among the elderly: the Framingham Study. Am J Epidemiol 2002;156:1021–7.
53. Sunk IG, Amoyo-Minar L, Niederreiter B, Soleiman A, Kainberger F, Smolen JS, *et al*. Histopathological correlation supports the use of X-rays in the diagnosis of hand osteoarthritis. Ann Rheum Dis 2013;72:572–7.
54. Grainger AJ, Farrant JM, O'Connor PJ, Tan AL, Tanner S, Emery P, *et al*. MR imaging of erosions in interphalangeal joint osteoarthritis: is all osteoarthritis erosive? Skeletal Radiol 2007;36:737–45.
55. Wittoek R, Jans L, Lambrecht V, Carron P, Verstraete K, Verbruggen G. Reliability and construct validity of ultrasonography of soft tissue and destructive changes in erosive

osteoarthritis of the interphalangeal finger joints: a comparison with MRI. Ann Rheum Dis 2011;70:278–83.
56. Iagnocco A, Filippucci E, Ossandon A, Ciapetti A, Salaffi F, Basili S, *et al*. High resolution ultrasonography in detection of bone erosions in patients with hand osteoarthritis. J Rheumatol 2005;32:2381–3.
57. Keen HI, Wakefield RJ, Grainger AJ, Hensor EM, Emery P, Conaghan PG. Can ultrasonography improve on radiographic assessment in osteoarthritis of the hands? A comparison between radiographic and ultrasonographic detected pathology. Ann Rheum Dis 2008;67:1116–20.
58. Vlychou M, Koutroumpas A, Malizos K, Sakkas LI. Ultrasonographic evidence of inflammation is frequent in hands of patients with erosive osteoarthritis. Osteoarthritis Cartilage 2009;17:1283–7.
59. Huetink K, van't Klooster R, Kaptein BL, Watt I, Kloppenburg M, Nelissen RG, *et al*. Automatic radiographic quantification of hand osteoarthritis; accuracy and sensitivity to change in joint space width in a phantom and cadaver study. Skelet Radiol 2012;41:41–9.
60. van't Klooster R, Hendriks EA, Watt I, Kloppenburg M, Reiber JH, Stoel BC. Automatic quantification of osteoarthritis in hand radiographs: validation of a new method to measure joint space width. Osteoarthritis Cartilage 2008;16:18–25.
61. Saltzherr MS, van Neck JW, Muradin GS, Ouwendijk R, Luime JJ, Coert JH, *et al*. Computed tomography for the detection of thumb base osteoarthritis: comparison with digital radiography. Skelet Radiol 2013;42:715–21.
62. Ceceli E, Gul S, Borman P, Uysal SR, Okumus M. Hand function in female patients with hand osteoarthritis: relation with radiological progression. Hand (NY) 2012;7:335–40.
63. Marshall M, van der Windt D, Nicholls E, Myers H, Hay E, Dziedzic K. Radiographic hand osteoarthritis: patterns and associations with hand pain and function in a community-dwelling sample. Osteoarthritis Cartilage 2009;17:1440–7.
64. Buckland-Wright JC, Macfarlane DG, Lynch JA. Sensitivity of radiographic features and specificity of scintigraphic imaging in hand osteoarthritis. Rev Rhum Engl Ed 1995;62:14S–26S.
65. Marshall M, Dziedzic KS, van der Windt DA, Hay EM. A systematic search and narrative review of radiographic definitions of hand osteoarthritis in population-based studies. Osteoarthritis Cartilage 2008;16:219–26.
66. Kerkhof HJ, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zeinstra SM, *et al*. Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage 2011;19:254–64.
67. Haugen IK, Boyesen P. Imaging modalities in hand osteoarthritis—and perspectives of conventional radiography, magnetic resonance imaging, and ultrasonography. Arthritis Res Ther 2011;13:248.
68. Spaans AJ, van Laarhoven CM, Schuurman AH, van Minnen LP. Interobserver agreement of the Eaton-Littler classification system and treatment strategy of thumb carpometacarpal joint osteoarthritis. J Hand Surg Am 2011;36:1467–70.
69. Sunk IG, Amoyo-Minar L, Stamm T, Haider S, Niederreiter B, Supp G, *et al*. Interphalangeal Osteoarthritis Radiographic Simplified (iOARS) score: a radiographic method to detect osteoarthritis of the interphalangeal finger joints based on its histopathological alterations. Ann Rheum Dis 2013 [epub ahead of print].
70. Altman RD, Fries JF, Bloch DA, Carstens J, Cooke TD, Genant H, *et al*. Radiographic assessment of progression in osteoarthritis. Arthritis Rheum 1987;30:1214–25.
71. Paradowski PT, Lohmander LS, Englund M. Natural history of radiographic features of hand osteoarthritis over 10 years. Osteoarthritis Cartilage 2010;18:917–22.
72. Maheu E, Cadet C, Carrat F, Barthe Y, Berenbaum F. Radiologic Progression of Hand Osteoarthritis Over 2.6 Years According to Various Methods of Calculation – Data From the SEKOIA Trial 2013.
73. Dela Rosa TL, Vance MC, Stern PJ. Radiographic optimization of the Eaton classification. J Hand Surg Br 2004;29:173–7.
74. Burnett S, Hart DJ, Cooper C, Spector TD. A Radiographic Atlas of Osteoarthritis 1994.
75. Eaton RG, Glickel SZ. Trapeziometacarpal osteoarthritis. Staging as a rationale for treatment. Hand Clin 1987;3:455–71.
76. Lane NE, Nevitt MC, Genant HK, Hochberg MC. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 1993;20:1911–8.