



**HAL**  
open science

## Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review

W.M. Reichmann, J.F. Maillefert, D.J. Hunter, J.N. Katz, P.G. Conaghan, E. Losina

### ► To cite this version:

W.M. Reichmann, J.F. Maillefert, D.J. Hunter, J.N. Katz, P.G. Conaghan, et al.. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review. *Osteoarthritis and Cartilage*, 2011, 19, pp.550-556. hal-00697253

**HAL Id: hal-00697253**

**<https://u-bourgogne.hal.science/hal-00697253>**

Submitted on 15 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review

W.M. Reichmann †, ‡, , ,

J.F. Maillefert §, ||, ¶,

D.J. Hunter#, † †,

J.N. Katz †, ‡ ‡, § §,

P.G. Conaghan|||,

E. Losina †, ‡

† Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA, USA

‡ Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

§ Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France

|| University of Burgundy, Dijon F-21079, France

¶ INSERM U887, Dijon F-21079, France

# Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW Australia

† † Division of Research, New England Baptist Hospital, Boston, MA, USA

‡ ‡ Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA, USA

§ § Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

||| Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK

Received 15 June 2010. Accepted 9 January 2011. Available online 9 March 2011.

<http://dx.doi.org/10.1016/j.joca.2011.01.023>, How to Cite or Link Using DOI

Cited by in Scopus (5)

Permissions & Reprints

---

Summary

Objective

The goal of this systematic review was to report the responsiveness to change and reliability of conventional radiographic joint space width (JSW) measurement.

## Method

We searched the PubMed and Embase databases using the following search criteria: [osteoarthritis (OA) (MeSH)] AND (knee) AND (X-ray OR radiography OR diagnostic imaging OR radiology OR disease progression) AND (joint space OR JSW or disease progression). We assessed responsiveness by calculating the standardized response mean (SRM). We assessed reliability using intra- and inter-reader intra-class correlation (ICC) and coefficient of variation (CV). Random-effects models were used to pool results from multiple studies. Results were stratified by study duration, design, techniques of obtaining radiographs, and measurement method.

## Results

We identified 998 articles using the search terms. Of these, 32 articles (43 estimates) reported data on responsiveness of JSW measurement and 24 (50 estimates) articles reported data on measures of reliability. The overall pooled SRM was 0.33 [95% confidence interval (CI): 0.26, 0.41].

Responsiveness of change in JSW measurement was improved substantially in studies of greater than 2 years duration (0.57). Further stratifying this result in studies of greater than 2 years duration, radiographs obtained with the knee in a flexed position yielded an SRM of 0.71. Pooled intra-reader ICC was estimated at 0.97 (95% CI: 0.92, 1.00) and the intra-reader CV estimated at 3.0 (95% CI: 2.0, 4.0). Pooled inter-reader ICC was estimated at 0.93 (95% CI: 0.86, 0.99) and the inter-reader CV estimated at 3.4% (95% CI: 1.3%, 5.5%).

## Conclusions

Measurement of JSW obtained from radiographs in persons with knee is reliable. These data will be useful to clinicians who are planning RCTs where the change in minimum JSW is the outcome of interest.

## Keywords

Knee osteoarthritis;

X-ray;

Radiograph;

Responsiveness;

Reliability;

Standardized response mean

---

## Introduction

Knee osteoarthritis (OA) is a painful and disabling disease for many with 12% of adults 60 years of age or older having symptomatic knee OA<sup>1</sup>. As the population ages, the prevalence of knee OA

continues to rise. Currently, available pharmacologic regimens for knee OA focus on alleviating pain, but do not slow the structural progression of disease<sup>2</sup>. Disease modifying OA drugs (DMOADS) are in the early developmental stages, and thus it is important to quantify the expected rate of structural progression to facilitate trial planning.

Minimum joint space width (JSW) is commonly used to assess knee OA progression<sup>3</sup>. It has been shown to be sensitive to change [4] and [5] and change in the minimum JSW has been the primary outcome for previous DMOAD trials [4], [5], [6] and [7]. An analytic literature synthesis by Emrani et al. in 2008 showed an interaction between study design and radiographic technique was associated with annual change in minimum JSW. The greatest annual change was seen in observational studies that used a semi-flexed technique without fluoroscopy, while the smallest annual change was seen in randomized controlled trials (RCTs) with the same technique<sup>5</sup>.

The objective of this paper was to update results of Emrani et al. by adding the most recent studies and report responsiveness of JSW in terms of standardized response mean (SRM). The SRM is defined as the mean change divided by the standard deviation (SD) of change and can be interpreted as the number of SDs of change, which will be useful for planning future DMOAD trials. We also report pooled estimates of reliability, which include inter- and intra-reader intra-class correlations (ICCs) and coefficients of variation (CVs).

## Method

### Eligibility criteria

Studies were eligible for our analyses if they satisfied all four requirements of the PICO (Patients Interventions Controls Outcomes). To be included in the review, the study population had to include patients with knee OA followed over time with radiograph-based measures of JSW. We included studies that reported responsiveness (mean change/SD of change or SRM) or reliability measures (inter- or intra-reader ICC or CV). If the study was a RCT then we used data from the control group. This was done to ensure quantification of the natural history of responsiveness of radiographs in those with knee OA. Studies were not limited by publication date (latest search: April 2009) and we included studies that were published in English, French, Spanish, and German.

### Information sources and search

We searched the PubMed and Embase databases using the following search criteria: (osteoarthritis [MeSH]) AND (knee) AND (X-ray OR radiography OR diagnostic imaging OR radiology OR disease progression) AND (joint space OR JSW or disease progression).

### Study selection

All abstracts were read by one reviewer. The reviewer obtained full-length articles of all abstracts that were considered as probably relevant or of unknown relevance. These articles were subsequently reviewed and data extracted into a data abstraction form. Abstracts of all potentially relevant references in the full-text review were obtained if probably relevant or of unknown relevance.

Studies were excluded if they did not report change in minimum JSW in the knee or if they did not provide a measure of reliability in measuring minimum JSW.

#### Data items

We abstracted the following study characteristics from each article: study design, radiographic technique, use of fluoroscopy, method of measurement, follow-up time, whether readers were blinded to the order of the radiographic studies, and sample size. Study design was classified as RCT or observational and radiographic technique was categorized as extended view or flexed (includes semi-flexed). Method of measuring minimum JSW was performed manually or using a computer. Follow-up time was categorized as 1-year or less, 1-2 years, or greater than 2 years.

#### Summary measures

The principal summary measure for our review is the SRM. In articles that reported the SRM directly, we abstracted the reported value. In articles that only reported mean change and SD of change, we calculated the SRM from the two reported measures. Inter- and intra-reader reliability measures (ICC, CV) were also abstracted from the articles.

#### Synthesis of results

Random-effects models were built to obtain pooled estimates for the SRM and reliability measures across studies adjusting for variability across the studies. Heterogeneity in the estimates was assessed using I-squared, which assesses the percentage of variation across studies that was due to between study variation.

Analyses were performed for all studies that reported these measures and by study characteristics, including study design, radiographic approach, radiographic technique, use of fluoroscopy, method of measurement, and follow-up time. Ninety-five percent CIs were derived for all estimates.

#### Results

##### Study selection

We identified 866 articles using our electronic search and another 132 were identified manually for a total of 998 articles. Two-hundred eighty-five articles met the initial abstract screening inclusion criteria and the full-text article was obtained and read for further screening. Of these, 32 articles reported responsiveness results (43 estimates) and 24 articles reported reliability results. Of the 24 articles reporting reliability results, the inter-reader ICC was reported eight times, the intra-reader ICC 17 times, the inter-reader CV six times, and the intra-reader CV 19 times (Fig. 1).

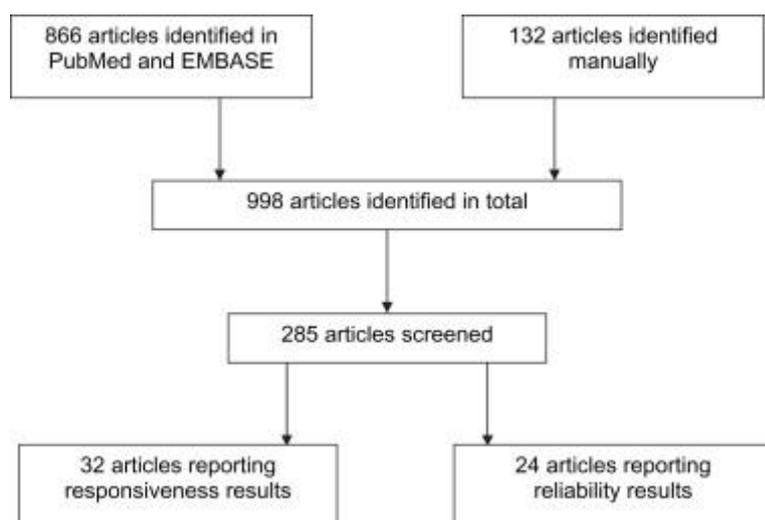


Fig. 1. Flow chart of the screening process for articles included in the systematic review.

[View thumbnail images](#)

[View high quality image \(139K\)](#)

[Study characteristics](#)

Of the 43 estimates on responsiveness, 21 (49%) estimates were obtained from studies with follow-up of 1 year or less, 10 (23%) estimates were derived from studies with follow-up of 1-2 years, and 12 (28%) came from studies with greater than 2 years of follow-up. The mean sample size was 100 (SD = 86). Sixteen estimates (37%) were obtained from studies that used a radiographic approach with the knee fully extended and 27 (63%) from studies that had the knee in flexion. Fluoroscopy was used for 23 (53%) of the estimates and computerized methods of measuring the minimum JSW was used for 24 of the estimates (56%). Nineteen (44%) of the estimates came from RCTs. Of the 43 estimates, only 21 (49%) disclosed whether the readers were blinded to the sequence of the radiographs. Of these 21 estimates, 19 (90%) came from studies that used blinded readers. Study characteristics for all 32 studies are shown in Table I.

Table I. Study Characteristics of the manuscripts reviewed for responsiveness

Author, year (Ref.)	Study type	Sample size	Follow-up months	Radiographic approach	Method of measurement	Delta (SD)
Ayral et al. 19968	Cohort	41	12	Extension without fluoroscopy	Manual	0.40 (1.00)
Ravaud et al. 19969	Cohort	55	12	Extension without fluoroscopy	Manual	0.42 (1.11)
Listrat et al. 199710	RCT	17	12	Extension without fluoroscopy	Manual	0.70 (1.20)
Pavelka et al.	RCT	139	60	Extension with	Manual	0.42

Author, year (Ref.)	Study type	Sample size	Follow-up months	Radiographic approach	Method of measurement	Delta (SD)
20004				fluoroscopy		(0.94)
Mazzuca et al. 20016	Cohort	402	31.60	Extension without fluoroscopy	Manual	0.37 (1.25)
Reginster et al. 20017	RCT	106	36	Extension with fluoroscopy	Computerized	0.40 (0.92)
Gandy et al. 200211	Cohort	11	37	Extension without fluoroscopy	Manual	0.21 (0.37)
Miyazaki et al. 200212	Cohort	74	72	Flexion without fluoroscopy	Manual	1.40 (1.20)
Boegard et al. 200313	Cohort	50	25	Flexion with fluoroscopy	Manual	0.06 (0.45)
Mazzuca et al. 200314	Cohort	52	14	Flexion with fluoroscopy	Computerized	0.09 (0.31)
		52	14	Flexion without fluoroscopy	Manual	-0.09 (0.66)
		20	12	Flexion with fluoroscopy	Manual	0.00 (0.60)
Pessis et al. 200315	Cohort	20	12	Extension with fluoroscopy	Manual	0.10 (0.90)
		20	12	Extension with fluoroscopy	Manual	0.10 (0.90)
Sugiyama et al. 200316	Cohort	110	48	Flexion with fluoroscopy	Computerized	0.53 (0.43)
Vignon et al. 200317	Cohort	58	24	Extension with fluoroscopy	Computerized	0.17 (0.75)
		58	24	Flexion with fluoroscopy	Computerized	0.24 (0.50)
Pavelka et al. 200418	RCT	89	24	Extension with fluoroscopy	Manual	0.40 (0.79)
Pham et al. 200419	RCT	79	12	Extension without fluoroscopy	Manual	0.21 (0.59)
		69	12	Extension without fluoroscopy	Manual	0.12

Author, year (Ref.)	Study type	Sample size	Follow-up months	Radiographic approach	Method of measurement	Delta (SD)
				fluoroscopy		(0.32)
Pham et al. 200420	RCT	277	12	Extension without fluoroscopy	Manual	0.09 (0.55)
Uebelhart et al. 200421	RCT	76	12	Extension without fluoroscopy	Computerized	0.32 (1.11)
Brandt et al. 200522	RCT	180	30	Flexion with fluoroscopy	Manual	0.45 (0.70)
Conrozier et al. 200523	Cohort	96	12	Flexion with fluoroscopy	Computerized	0.19 (0.48)
Michel et al. 200524	RCT	150	24	Flexion without fluoroscopy	Computerized	0.07 (0.56)
Spector et al. 200525	RCT	98	12	Flexion with fluoroscopy	Computerized	0.12 (0.42)
Bingham et al. 200626	RCT	269	24	Flexion with fluoroscopy	Computerized	0.13 (1.08)
		280	24	Flexion with fluoroscopy	Computerized	0.09 (1.31)
		112	9.84	Flexion without fluoroscopy	Computerized	0.00 (0.53)
Cline et al. 200627	RCT	85	11.76	Flexion with fluoroscopy	Computerized	0.12 (0.42)
		99	8.16	Flexion without fluoroscopy	Computerized	-0.07 (0.63)
Mikesky et al. 200628	RCT	60	30	Flexion with fluoroscopy	Manual	0.54 (0.70)
Botha-Scheepers et al. 200729	Cohort	122	24	Flexion without fluoroscopy	Computerized	0.21 (0.52)
Krzeski et al. 200730	RCT	71	12	Extension with fluoroscopy	N/A	0.14 (0.53)
Nevitt et al.	Cohort	53	37	Flexion without	Computerized	0.43

Author, year (Ref.)	Study type	Sample size	Follow-up months	Radiographic approach	Method of measurement	Delta (SD)
200731				fluoroscopy		(0.66)
Sharif et al. 200732	Cohort	115	60	Extension without fluoroscopy	Manual	0.18 (0.93)
Le Graverand et al. 200833	Cohort	62	12	Flexion with fluoroscopy	Computerized	0.22 (0.41)
		62	12	Flexion without fluoroscopy	Computerized	-0.01 (0.46)
		27	12	Flexion without fluoroscopy	Computerized	0.25 (0.54)
Mazzuca et al. 200834	Cohort	27	12	Flexion without fluoroscopy	Computerized	0.02 (0.40)
		47	12	Flexion with fluoroscopy	Computerized	0.16 (0.37)
		47	12	Flexion with fluoroscopy	Computerized	-0.01 (0.51)
Gensburger et al. 200935	Cohort	81	48	Flexion with fluoroscopy	Computerized	0.32 (0.76)
Kahan et al. 200936	RCT	313	12	Flexion with fluoroscopy	Computerized	0.31 (0.71)

Delta: change in minimum JSW from baseline to follow-up (measured in millimeters).

#### Full-size table

Of the eight estimates evaluating the inter-reader ICC, four (50%) used a fully extended radiographic approach, four (50%) used fluoroscopy, and seven (88%) measured the joint space manually. The mean sample size in these studies was 110 (SD = 110).

Of the 17 estimates evaluating the intra-reader ICC, six (35%) used a fully extended radiographic approach, eight (47%) used fluoroscopy, and nine (53%) measured the joint space manually. The mean sample size in these studies was 80 (SD = 88).

Of the six estimates evaluating the inter-reader CV, three (50%) used a fully extended radiographic approach, six (100%) used fluoroscopy, and six (100%) measured the joint space manually. The mean sample size in these studies was 120 (SD = 99).

Of the 19 estimates evaluating the intra-reader CV, six (32%) used a fully extended radiographic approach, 14 (74%) used fluoroscopy, and 11 (58%) measured the joint space manually. The mean sample size was 43 (SD = 38). Study characteristics for all 24 studies are shown in Table II.

Table II. Study characteristics of the manuscripts reviewed for reliability

Author, year (Ref.)	Sample size	Radiographic approach	Method of measurement	Reliability estimator	Observer	Value
Buckland-Wright et al. 1995 <sup>37</sup>	5	Flexion with fluoroscopy	Computer	CV	Intra	3.8%
	5	Flexion with fluoroscopy	Computer	CV	Intra	1.2%
	7	Flexion with fluoroscopy	Manual	CV	Intra	3.6%
	7	Flexion with fluoroscopy	Manual	CV	Intra	0.6%
Ravaud et al. 1996 <sup>9</sup>	55	Extension without fluoroscopy	Manual	ICC	Intra	0.95
	55	Extension without fluoroscopy	Manual	ICC	Inter	0.85
Pavelka et al. 2000 <sup>4</sup>	10	Extension with fluoroscopy	Manual	CV	Intra	2.0%
	10	Extension with fluoroscopy	Manual	ICC	Intra	0.99
	280	Extension with fluoroscopy	Manual	CV	Inter	6.6%
	280	Extension with fluoroscopy	Manual	ICC	Inter	0.97
Mazucca et al. 2001 <sup>6</sup>	20	Extension without fluoroscopy	Manual	CV	Intra	4.4%
Myazaki et al. 2002 <sup>12</sup>	10	Flexion without fluoroscopy	Manual	ICC	Intra	0.92
Pavelka et al. 2002 <sup>38</sup>	40	Extension with fluoroscopy	Manual	CV	Intra	1.9%

Author, year (Ref.)	Sample size	Radiographic approach	Method of measurement	Reliability estimator	Observer	Value
Boegard et al. 200313	202	Extension with fluoroscopy	Manual	CV	Inter	2.6%
	51	Flexion with fluoroscopy	Manual	CV	Intra	2.3%
	51	Flexion with fluoroscopy	Manual	CV	Intra	1.0%
	51	Flexion with fluoroscopy	Manual	CV	Inter	2.7%
	51	Flexion with fluoroscopy	Manual	CV	Inter	1.1%
Mazzuca et al. 200314	71	Flexion without fluoroscopy	Manual	CV	Intra	5.8%
Sugiyama et al. 200316	10	Flexion with fluoroscopy	Computer	CV	Intra	1.5%
Vignon et al. 200317	20	Extension with fluoroscopy	Computer	ICC	Intra	0.98
	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.98
Mazzuca et al. 200439	30	Flexion with fluoroscopy	Manual	ICC	Intra	0.996
	30	Flexion with fluoroscopy	Manual	ICC	Inter	0.956
Pavelka et al. 200418	89	Extension with fluoroscopy	Manual	CV	Intra	3.6%
	89	Extension with fluoroscopy	Manual	CV	Inter	6.5%
Pham et al. 200419	156	Extension without fluoroscopy	Manual	ICC	Intra	0.993
Pham et al. 200420	292	Extension without fluoroscopy	Manual	ICC	Intra	0.996

Author, year (Ref.)	Sample size	Radiographic approach	Method of measurement	Reliability estimator	Observer	Value
	292	Extension without fluoroscopy	Manual	ICC	Inter	0.912
Sharif et al. 200440	20	Extension without fluoroscopy	Manual	CV	Intra	11.3%
Cicuttini et al. 200541	123	Extension without fluoroscopy	Computer	CV	Intra	4.8%
Conrozier et al. 200523	106	Flexion with fluoroscopy	Computer	CV	Intra	1.15%
	106	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
Michel et al. 200524	284	Flexion without fluoroscopy	Computer	ICC	Intra	0.98
Szebenyi et al. 200642	60	Extension without fluoroscopy	Manual	ICC	Intra	0.895
	60	Extension without fluoroscopy	Manual	ICC	Inter	0.868
	80	Flexion without fluoroscopy	Manual	ICC	Intra	0.90
Nevitt et al. 200731	80	Flexion without fluoroscopy	Manual	ICC	Inter	0.98
	25	Flexion without fluoroscopy	Computer	ICC	Intra	0.96
	25	Flexion without fluoroscopy	Computer	CV	Intra	2.9%
	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
Le Graverand et al. 200833	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
	18	Flexion without fluoroscopy	Computer	ICC	Intra	0.99

Author, year (Ref.)	Sample size	Radiographic approach	Method of measurement	Reliability estimator	Observer	Value
Mazzuca et al. 200834	39	Flexion with fluoroscopy	Computer	CV	Intra	0.80
	42	Flexion with fluoroscopy	Manual	ICC	Intra	0.89
Gensburger et al. 200935	42	Flexion with fluoroscopy	Manual	CV	Intra	2.9%
	44	Flexion with fluoroscopy	Manual	ICC	Inter	0.80
	44	Flexion with fluoroscopy	Manual	CV	Inter	0.8%
Kahan et al. 200936	100	Flexion with fluoroscopy	Computer	CV	Intra	1.2%
	100	Flexion with fluoroscopy	Computer	ICC	Intra	0.99

Full-size table

#### Synthesis of responsiveness results

The I-squared value for the 43 estimates was 0.82 [95% confidence interval (CI): 0.76, 0.86] indicating substantial between study variation. The I-squared values are shown in Table III.

Table III. Results of random-effects pooling for studies that reported estimates of responsiveness by different study characteristics

	Number of estimates	I-squared (95% CI)	SRM (95% CI)
Overall	43	0.82 (0.76, 0.86)	0.33 (0.26, 0.41)
Knee flexion			
Extended	16	0.19 (0.00, 0.55)	0.32 (0.26, 0.37)
Flexed	27	0.88 (0.84, 0.91)	0.34 (0.22, 0.45)
Fluoroscopy			
Fluoro	23	0.83 (0.76, 0.88)	0.38 (0.27, 0.48)

	Number of estimates	I-squared (95% CI)	SRM (95% CI)
No fluoro	20	0.79 (0.69, 0.86)	0.28 (0.17, 0.39)
Measurement method			
Manual	18	0.80 (0.70, 0.87)	0.38 (0.26, 0.50)
Computerized	24	0.84 (0.77, 0.89)	0.31 (0.20, 0.41)
Study type			
RCT	19	0.82 (0.73, 0.88)	0.30 (0.20, 0.40)
Cohort	24	0.82 (0.74, 0.87)	0.36 (0.24, 0.49)
Follow-up time			
1-year or less	21	0.56 (0.27, 0.73)	0.24 (0.15, 0.32)
1-2 years	10	0.80 (0.63, 0.89)	0.25 (0.13, 0.37)
Greater than 2 years	12	0.88 (0.81, 0.93)	0.57 (0.39, 0.75)
Reader blinded to order of radiographs			
Yes	19	0.76 (0.63, 0.85)	0.30 (0.19, 0.40)
No	2	0.59 (0.00, 0.90)	0.55 (0.33, 0.76)
Unknown	22	0.85 (0.78, 0.89)	0.35 (0.23, 0.46)
Knee flexion by follow-up time			
Extended/1-year or less	9	0.00 (0.00, 0.63)	0.26 (0.19, 0.34)
Extended/1-2 years	2	0.61 (0.00, 0.91)	0.38 (0.10, 0.65)
Extended/greater than 2 years	5	0.32 (0.00, 0.74)	0.34 (0.24, 0.44)
Flexed/1-year or less	12	0.68 (0.42, 0.83)	0.19 (0.06, 0.32)
Flexed/1-2 years	8	0.82 (0.65, 0.90)	0.22 (0.08, 0.36)
Flexed/greater than 2 years	7	0.88 (0.78, 0.94)	0.71 (0.44, 0.98)
Fluoroscopy by follow-up time			
Fluoro/1-year or less	9	0.33 (0.00, 0.69)	0.29 (0.18, 0.39)

	Number of estimates	I-squared (95% CI)	SRM (95% CI)
Fluoro/1-2 years	7	0.81 (0.62, 0.91)	0.29 (0.14, 0.44)
Fluoro/greater than 2 years	7	0.87 (0.75, 0.93)	0.58 (0.36, 0.80)
No fluoro/1-year or less	12	0.61 (0.28, 0.79)	0.21 (0.10, 0.32)
No fluoro/1-2 years	3	0.82 (0.45, 0.94)	0.15 (-0.13, 0.42)
No fluoro/greater than 2 years	5	0.89 (0.78, 0.95)	0.56 (0.24, 0.87)
Measurement method by follow-up time			
Manual/1-year or less	8	0.20 (0.00, 0.63)	0.28 (0.17, 0.38)
Manual/1-2 years	2	0.92 (0.73, 0.98)	0.19 (-0.44, 0.82)
Manual/greater than 2 years	8	0.87 (0.77, 0.93)	0.51 (0.31, 0.71)
Computerized/1-year or less	12	0.68 (0.42, 0.83)	0.21 (0.08, 0.33)
Computerized/1-2 years	8	0.78 (0.56, 0.89)	0.26 (0.13, 0.38)
Computerized/greater than 2 years	4	0.90 (0.77, 0.96)	0.68 (0.31, 1.06)
Study type by follow-up time			
RCT/1-year or less	10	0.60 (0.19, 0.80)	0.21 (0.11, 0.32)
RCT/1-2 years	5	0.87 (0.72, 0.94)	0.24 (0.07, 0.41)
RCT/greater than 2 years	4	0.51 (0.00, 0.84)	0.56 (0.41, 0.70)
Cohort/1-year or less	11	0.51 (0.03, 0.75)	0.26 (0.13, 0.40)
Cohort/1-2 years	5	0.69 (0.20, 0.88)	0.26 (0.06, 0.46)
Cohort/greater than 2 years	8	0.92 (0.86, 0.95)	0.57 (0.30, 0.85)

#### Full-size table

The random-effects analysis yielded an overall pooled SRM for the 43 estimates of 0.33 (95% CI: 0.26, 0.41). The pooled SRM was similar when the analysis was stratified by radiographic approach, the use of fluoroscopy, measurement method, and study type. Follow-up time was related to the magnitude of the SRM. Estimates derived from studies with 1 year or less and 1-2 years of follow-up had similar

responsiveness (0.24 and 0.25 respectively), while estimates coming from studies with greater than 2 years follow-up had an SRM of 0.57 (95% CI: 0.39, 0.75). Similar effects of follow-up time are shown when use of fluoroscopy, measurement method, and study type were stratified by follow-up time. However, when radiographic approach was stratified by follow-up time, estimates derived from studies that used a flexion-based radiographic approach and had greater than 2 years of follow-up time had a higher SRM of 0.71 (95% CI: 0.44, 0.98).

#### Synthesis of reliability results

Results of random-effects pooling of the reliability estimates showed good inter- and intra-reader reliability for measuring the minimum JSW. The eight estimates of inter-reader ICC produced an estimate of 0.93 (95% CI: 0.86, 0.99), while the 17 estimates of intra-reader ICC produced an estimate of 0.97 (95% CI: 0.92, 1.00). Additional results stratified by study characteristics are shown in Table IV. Six estimates for the inter-reader CV produced an estimate of 3.4% (95% CI: 1.3%, 5.5%) and 19 estimates for the intra-reader CV produced an estimate of 3.0% (95% CI: 2.0%, 4.0%). Additional results stratified by study characteristics are shown in Table V.

Table IV. Results of random-effects pooling for studies that reported estimates of intra-ICC by different study characteristics

	Number of estimates	Inter-reader ICC (95% CI)	Number of estimates	Intra-reader ICC (95% CI)
Overall	8	0.93 (0.86, 0.99)	17	0.97 (0.92, 1.00)
Knee flexion				
Extended	4	0.93 (0.85, 1.00)	6	0.98 (0.90, 1.00)
Flexed	4	0.94 (0.79, 1.00)	11	0.97 (0.90, 1.00)
Fluoroscopy				
Fluoro	4	0.95 (0.85, 1.00)	8	0.98 (0.88, 1.00)
No fluoro	4	0.91 (0.82, 1.00)	9	0.97 (0.91, 1.00)
Measurement method				
Manual	7	0.93 (0.86, 0.99)	9	0.97 (0.89, 1.00)
Computerized	1	0.99 (N/A)	8	0.98 (0.90, 1.00)

Full-size table

Table V. Results of random-effects pooling for studies that reported estimates of CV by different study characteristics

	Number of estimates	Inter-reader CV (95% CI)	Number of estimates	Intra-reader CV (95% CI)
Overall	6	3.4% (1.3, 5.5)	19	3.0% (2.0, 4.0)
Knee flexion				
Extended	3	5.2% (2.5, 8.0)	6	4.7% (2.7, 6.7)
Flexed	3	1.5% (0.3, 2.7)	13	2.2% (1.3, 3.2)
Fluoroscopy				
Fluoro	6	3.4% (1.3, 5.5)	14	2.0% (1.4, 2.5)
No fluoro	0	N/A	5	5.8% (3.8, 7.9)
Measurement method				
Manual	6	3.4% (1.3, 5.5)	11	3.6% (2.1, 5.1)
Computerized	0	N/A	8	2.2% (0.8, 3.5)

Full-size table

## Discussion

We performed an analytic systematic review of the responsiveness and reliability of knee radiographs when measuring the minimum JSW. We analyzed responsiveness using the SRM. This measure can be interpreted as the number of SDs of change. The overall SRM was 0.33 (95% CI: 0.26, 0.41). Follow-up time was the main study characteristic that was related to responsiveness. Studies with follow-up times greater than 2 years showed greater responsiveness (SRM = 0.57; 95% CI: 0.39, 0.75). It is critical to note that studies with a follow-up of 1 year or shorter showed a responsiveness of 0.24. This limitation of the radiographic technique means that to adequately power a study to demonstrate change over this short interval will require much larger sample sizes. Studies that used a flexed view and had greater than 2 years of follow-up showed the greatest responsiveness (SRM = 0.71; 95% CI: 0.44, 0.98). Based upon this literature there does appear to be some advantage to standardized positioning and fluoroscopy with slight improvements in responsiveness. Despite what one may have expected there does not appear to be any advantage in computerized measurement of JSW over manual measures. In studies with greater than 2 years of follow-up, the responsiveness was higher for those that used computerized methods (0.68) compared to those that used manual methods (0.51). However, the 95% CIs substantially overlap due to substantial variability in these estimates (Table III).

The reliability of measuring minimum JSW provided to be excellent with pooled ICCs ranging from 0.91 to 0.99 and pooled CVs ranging from 1.5 to 5.8. Radiographic method, use of fluoroscopy, and measurement method did not affect reliability albeit the majority of the estimates come from different studies with no direct study comparison.

Our findings complement the work of Emrani et al. who published a systematic review in 2008 on the change in minimum JSW. While they found effects of radiographic approach and study type, they also analyzed the crude change in minimum JSW rather than the SRM. They also found that increased follow-up time was inversely associated with change in minimum JSW, while we found that increasing the follow-up time increased the responsiveness of radiographs to change. This difference may be due to differences in definition of primary outcomes and additional assumption of linearity of change that Emrani et al. used in their analysis<sup>5</sup>.

A major strength of this study is that it is the first literature synthesis to summarize responsiveness in terms of the SRM. These data will be useful to clinicians who are planning studies where the change in the minimum JSW is the outcome of interest. The results of this analysis suggest that studies using JSW as primary outcome measure based on radiographs should plan to have a follow-up period that is greater than 2 years and have the knee in a flexed position when performing the radiographs to ensure the greatest possible responsiveness. While the pooled SRM was higher for studies that did not blind the reader to the sequence of the radiographs (0.55), it is unlikely that blinding of the readers of the radiographs substantially influenced our results since only two estimates came from studies that did not blind their readers. Also, the pooled SRM for estimates coming from studies that did blind the readers was similar to those that did not report this information (0.30 vs 0.35 respectively).

Also, this is the first known literature synthesis that pools reliability data on measuring minimum JSW. In general, these measurements can be considered to be reliable as the intra- and inter-reader ICCs were large and the CVs were low.

A major limitation of our review is that we did not report our results by risk factors for knee OA progression (body mass index, knee alignment, age, concurrent OA in other joints, synovitis, etc.) since they were not uniformly reported. The fact that we were not able to account for these factors may have contributed to the heterogeneity in the SRMs. It is important for future studies that report results on quantitative changes of knee OA progression to report these risk factors. Also, we did not collect data on the number of readers and the time interval between reads for our reliability data. It would be interesting to examine how these factors affected our estimates of reliability.

We found that radiographs provide moderate responsiveness and good reliability measures for measuring the minimum JSW in persons with knee OA. These data will be useful to clinicians who wish to plan future RCTs in which change in minimum JSW is their primary outcome.

Declaration of funding and role of funding source

The Osteoarthritis Research Society International (OARSI) Food and Drug Administration (FDA) OA Initiative received financial support from the following professional organization: American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab, AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew, Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA. This work was also supported in part by the National Institutes of Health (K24 AR057827 and T32 AR055885) and they had no role in the content being published.

#### Author contributions

.

Conception and design (WMR, JFM, EL).

.

Analysis and interpretation of the data (WMR, JFM, DJH, PGC, JNK, EL).

.

Drafting of the article (WMR, JFM, DJH, PGC, JNK, EL).

.

Critical revision of the article for important intellectual content (WMR, JFM, DJH, PGC, JNK, EL).

.

Final approval of the article (WMR, JFM, DJH, PGC, JNK, EL).

.

Statistical expertise (WMR, EL).

.

Collection and assembly of data (WMR, JFM).

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.