

## EXTENDED REPORT

# Validity and reliability of three definitions of hip osteoarthritis: cross sectional and longitudinal approach

M Reijman, J M W Hazes, H A P Pols, R M D Bernsen, B W Koes, S M A Bierma-Zeinstra

*Ann Rheum Dis* 2004;63:1427–1433. doi: 10.1136/ard.2003.016477

See end of article for authors' affiliations

Correspondence to:  
Dr M Reijman, Department of General Practice, Erasmus MC – Faculty, PO Box 1738, 3000 DR Rotterdam, Netherlands; m.reijman@erasmusmc.nl

Accepted  
21 December 2003

**Objectives:** To compare the reliability and validity in a large open population of three frequently used radiological definitions of hip osteoarthritis (OA): Kellgren and Lawrence grade, minimal joint space (MJS), and Croft grade; and to investigate whether the validity of the three definitions of hip OA is sex dependent.

**Methods:** Subjects from the Rotterdam study (aged  $\geq 55$  years,  $n = 3585$ ) were evaluated. The inter-rater reliability was tested in a random set of 148 x rays. The validity was expressed as the ability to identify patients who show clinical symptoms of hip OA (construct validity) and as the ability to predict total hip replacement (THR) at follow up (predictive validity).

**Results:** Inter-rater reliability was similar for the Kellgren and Lawrence grade and MJS ( $\kappa$  statistics 0.68 and 0.62, respectively) but lower for Croft's grade ( $\kappa$  statistic, 0.51). The Kellgren and Lawrence grade and MJS showed the strongest associations with clinical symptoms of hip OA. Sex appeared to be an effect modifier for Kellgren and Lawrence and MJS definitions, women showing a stronger association between grading and symptoms than men. However, the sex dependency was attributed to differences in height between women and men. The Kellgren and Lawrence grade showed the highest predictive value for THR at follow up.

**Conclusions:** Based on these findings, Kellgren and Lawrence still appears to be a useful OA definition for epidemiological studies focusing on the presence of hip OA.

Osteoarthritis (OA) of the hip is of particular interest as this is often the sole joint affected by the disease, suggesting an important role for local biomechanical risk factors. In addition, the prevalence of hip OA is expected to increase with the aging of Western society,<sup>1</sup> and the hip is crucial for independent function.<sup>2</sup>

A problem with studying hip OA is the absence of a consensus for defining it for the purposes of epidemiological research.<sup>3</sup> To investigate occurrence and (potential) risk factors, a valid and reliable definition of hip OA is required. Most epidemiological studies have used a single hallmark (radiological signs) to define hip OA.<sup>4–5</sup>

In a previous systematic appraisal, we summarised the validity, reliability, and applicability of seven definitions of hip OA used in epidemiological studies.<sup>6</sup> Considering the frequent use of the definitions of hip OA, it is surprising that the validity of these definitions has been so poorly investigated. Because of the lack of comparability between the different studies and because most studies only investigated a single definition, it was difficult to compare the reliability and validity of the various definitions of hip OA. Our appraisal also showed that the validity and reliability of minimal joint space (MJS; according to Croft) and Croft's grade (a modification of the Kellgren and Lawrence grade)<sup>7</sup> have only been studied in a male population.

The primary objective of the present study was to compare the reliability and validity of the three most commonly used radiological definitions of hip OA—Kellgren and Lawrence grade, MJS (according to Croft), and Croft's grade—in a large open population of elderly people. Our secondary objective was to investigate whether the validity of the three definitions of hip OA was sex dependent.

## METHODS

The study population consisted of participants in the Rotterdam study, a prospective cohort of men and women

aged 55 years and over. The objective of the Rotterdam study is to investigate the incidence of and risk factors for chronic disabling diseases. The rationale and study design have been described previously.<sup>8</sup> The focus is on neurogeriatric, cardiovascular, ophthalmological, and locomotor diseases. All 10 275 inhabitants of Ommoord (a district in Rotterdam, Netherlands) were invited to participate. The response rate was 78%, resulting in 7983 subjects participating in the study. Written informed consent was obtained from each participant. The medical ethics committee of the Erasmus University Medical Centre approved the Rotterdam Study.

We used a sample of 3585 subjects from the Rotterdam study. Selection was based on the availability of the radiographs of the hip at baseline and follow up. The fact that subjects had to be mobile enough to visit the research centre at baseline and follow up and survive the follow up period caused a health selection bias in our study population. Compared with the total Rotterdam study population, the study population was significantly younger (66.0 years *v* 70.6 years), had a lower prevalence of lower limb disability at baseline (index score  $\geq 0.5$ : 12.9% *v* 35.5%), and had a somewhat lower prevalence of hip pain at baseline (11.7% *v* 12.7%).

Subjects with bilateral total hip replacement (THR) at baseline ( $n = 24$ ) were excluded from analysis, which resulted in a study population of 3561 subjects. The baseline measurements were conducted between April 1990 and July 1993, and the follow up measurements between 1996 and 1999, with a mean (SD) follow up time of 6.6 (0.50) years.

**Abbreviations:** GEE, generalised estimating equations; HAQ, health assessment questionnaire; ICC, intraclass correlation coefficient; LDLI, lower limb disability index; MJS, minimal joint space; ROA, radiographic osteoarthritis; THR, total hip replacement

**Table 1** Demographic characteristics and prevalences of radiographic hip osteoarthritis stratified for sex

Variable	Men (n = 1499)	Women (n = 2086)	Total (n = 3585)
Age (years) (mean (SD))	65.5 (6.5)	66.3 (7.2)	66.0 (6.9)
BMI (kg/m <sup>2</sup> ) (mean (SD))	25.9 (2.8)	26.6 (4.0)	26.3 (3.6)
Height (cm) (mean (SD))	175.5 (6.6)	162.4 (6.5)	167.9 (9.2)
Lower limb disability ( $\geq 0.5$ ) (%)	8.2	16.3	12.9
Lower limb disability ( $\geq 1.0$ ) (%)	4.7	9.6	7.6
Morning stiffness (%)	24.9	36.9	31.9
Hip pain (left and/or right) (%)	7.1	14.9	11.7
Right	5.6	11.0	8.7
Left	4.5	9.9	7.6
Total hip replacement at baseline			
Unilateral (n (%))	20 (1.0)	69 (2.6)	89 (2.0)
Bilateral (n (%))	6 (0.4)	18 (0.9)	24 (0.7)
Kellgren (%)			
$\geq$ grade 2	7.8	6.4	7.0
$\geq$ grade 3	1.2	1.5	1.4
MJS (%)			
$\leq 2.5$ mm	6.8	8.1	7.5
$\leq 2.0$ mm	2.6	3.3	3.0
$\leq 1.5$ mm	1.2	1.5	1.4
Croft grade (%)			
$\geq$ grade 3	39.1	30.0	33.9
$\geq$ grade 4	4.5	4.4	4.4

BMI, body mass index; MJS, minimal joint space.

**Radiographic assessment**

Weight bearing anteroposterior pelvic radiographs with both feet in 10° internal rotation were obtained at 70 kV, a focus of 1.8, and a focus to film distance of 120 cm, applying a Fuji High Resolution G 35×43 cm film.<sup>9</sup> The x ray beam was centred on the umbilicus.

One independent trained reader (MR) evaluated the radiographs according to a standardised protocol, unaware of the clinical status of the patients.

At baseline, radiographic osteoarthritis (ROA) of the hip was quantified by measurements of the Kellgren and Lawrence grading system (atlas based) (table 7),<sup>6 10-13</sup> the Croft grading system (a modification of the Kellgren and Lawrence system) (table 8), and MJS as defined by Croft (table 9).<sup>6 7 12-14</sup> For the Croft grading scale, we assessed the individual radiographic features of MJS, presence of osteophytes, subchondral sclerosis, and cyst formation. We looked

for the presence of the individual radiographic features (of any grade), using an atlas of individual features.<sup>12 13</sup> Different cut off points to quantify hip ROA were employed: for Kellgren and Lawrence,  $\geq$ grade 2 (moderate) and  $\geq$ grade 3 (severe); for the Croft grading system,  $\geq$ grade 3 (moderate) and  $\geq$ grade 4 (severe); and for MJS,  $\leq 2.5$  mm (moderate),  $\leq 2.0$  mm (intermediate), and  $\leq 1.5$  mm (severe).

The joint space width (lateral, superior, axial, medial, and minimum) measurements were standardised using a 0.5 mm graduated magnifying glass laid directly over the radiograph.<sup>15-17</sup>

The follow up radiographs were evaluated for the presence of an incident THR (not present at baseline).

For all three grading systems and all measurements, inter-rater reliability was tested in a random set of 148 radiographs.<sup>18 19</sup>

**Clinical assessment**

At baseline, trained interviewers undertook an extensive home interview on demographic characteristics, medical history, risk factors for chronic diseases, and use of medicines.

For this study we used information on the presence of hip pain (“Did you have joint complaints of your right/left hip during the last month”), the presence of morning stiffness, and lower limb disability. Lower limb disability was assessed using a modified version of the Stanford health assessment questionnaire.<sup>9</sup> A lower limb disability index (LDI) was obtained by calculating the mean score of answers to the following six questions: “Are you able to stand up from a straight chair without using your arms for support?”; “Are you able to get in and out of bed?”; “Are you able to walk outdoors on flat ground?”; “Are you able to climb up five steps?”; “Are you able to bend down to pick up clothing from the floor?”; and “Are you able to get in and out of a car?”. The answers were scored as follows: 0 = yes, without difficulty, 1 = yes, with some difficulty, 2 = yes, with much difficulty, 3 = no, unable to do (needs help). Moderate disability was defined as a score greater than 0.5 and severe disability as a score greater than 1.0 on the LDI. Moderate disability is present whenever there is at least some difficulty with three of six daily activities in the LDI.<sup>9</sup>

**Table 2** Inter-rater reliability for individual radiological features and the three definitions of radiographic hip osteoarthritis studied (n = 148)

Definition of OA	Inter-rater reliability
<i>Radiological features</i>	
<i>Subchondral sclerosis</i>	
Acetabulum	0.51 (0.35 to 0.67)
Femoral head	0.66 (0.22 to 1.00)
<i>Osteophytes</i>	
Acetabulum	0.23 (0.06 to 0.40)
Femoral head	0.34 (0.19 to 0.50)
Cysts	0.32 (-0.17 to 0.82)
MJS (continuous)	0.85* (0.80 to 0.89)
<i>Definitions studied</i>	
Kellgren, $\geq$ grade 2	0.68 (0.44 to 0.92)
MJS, $\leq 2.5$ mm	0.62 (0.43 to 0.81)
Croft grade, $\geq$ grade 3	0.51 (0.35 to 0.67)

Inter-rater reliability is presented by intraclass correlation coefficient (two way mixed effect model, consistency definition) for minimal joint space (as continuous variable) and by  $\kappa$  for other individual radiological features and three definitions of hip OA studied (95% confidence intervals in parentheses). MJS, minimal joint space.

**Table 3** Association between different definitions of radiographic hip osteoarthritis and clinical symptoms (n = 3561)

	Hip pain		Morning stiffness OR (95% CI)	Disability (LLD $\geq$ 0.5) OR (95% CI)	Disability (LLD $\geq$ 1.0) OR (95% CI)
	OR (95% CI)	Sensitivity/specificity (%)			
<b>Kellgren</b>					
$\geq$ grade 2	2.6 (1.8 to 3.6)	20.7/92.7	1.2 (0.9 to 1.6)	2.4 (1.7 to 3.4)	3.0 (2.0 to 4.4)
$\geq$ grade 3	6.6 (3.6 to 12.1)	12.5/97.9	2.2 (1.2 to 3.9)	7.0 (3.8 to 12.8)	7.6 (4.0 to 14.6)
<b>MJS</b>					
$\leq$ 2.5 mm	2.4 (1.7 to 3.4)	14.9/93.3	1.6 (1.2 to 2.1)	2.7 (2.0 to 3.7)	3.0 (2.0 to 4.4)
$\leq$ 2.0 mm	4.5 (2.9 to 7.0)	9.3/97.8	1.7 (1.2 to 2.6)	3.7 (2.4 to 5.9)	4.1 (2.5 to 7.0)
$\leq$ 1.5 mm	6.6 (3.6 to 12.2)	5.5/99.1	2.0 (1.1 to 3.7)	5.3 (2.9 to 9.8)	6.1 (3.1 to 12.1)
<b>Croft grade</b>					
$\geq$ grade 3	1.3 (1.1 to 1.7)	39.9/66.9	0.9 (0.7 to 1.0)	1.0 (0.8 to 1.3)	1.1 (0.8 to 1.5)
$\geq$ grade 4	3.6 (2.4 to 5.2)	11.7/96.4	1.6 (1.1 to 2.2)	3.3 (2.3 to 4.9)	3.7 (2.3 to 5.8)

Associations are presented as odds ratios (OR) with 95% confidence intervals (CI) in parentheses. Associations are adjusted for body mass index and radiographic osteoarthritis of the other hip. LLD, lower limb disability index (a score of  $\geq$ 0.5 is defined as moderately disabled; a score of  $\geq$ 1.0 as severe disabled); MJS, minimal joint space.

**Statistical analysis**

For the inter-rater reliability, the  $\kappa$  statistic and the intraclass correlation coefficient (ICC) were assessed for the individual different radiological features and the three definitions of hip ROA.

Because of the absence of a gold standard, we expressed validity as *construct validity* and *predictive validity*. Construct validity is a measure of the ability to identify patients with symptoms (presence of hip pain, morning stiffness, or lower limb disability) of hip OA<sup>20,21</sup>; predictive validity is a measure of the ability to predict important long term outcomes of disease.<sup>21</sup> For construct validity, we tested the association between baseline radiological osteoarthritis of the hip according to the three definitions and the separate baseline clinical symptoms (hip pain, morning stiffness, and lower limb disability) by means of generalised estimating equations (GEE) (cross sectional design). This is a procedure of repeated measurements. It is used here to take account of the correlation between the left and right hip. In addition, sensitivity and specificity were assessed using the main symptom of hip OA—hip pain—as the gold standard. We used different cut off points for the three definitions of hip ROA and also stratified the results for sex and age. A two sided probability (p) value of 0.05 was considered significant.

For predictive validity, we assessed the proportion of THR after a clinically meaningful follow up period of 6.6 years in patients identified by each definition as having hip ROA at baseline (longitudinal design). We also calculated the association between the different definitions of hip ROA and THR at follow up by means of the GEE method (odds ratios). We used SPSS version 11.0 (SPSS Inc, Chicago, Illinois, USA) and SAS software, version 8.0 (SAS Institute, Cary, North Carolina, USA) for all analyses.

**RESULTS**

**Study population**

Table 1 shows the demographic characteristics and prevalence data on radiographic hip OA, stratified for sex of the study population of 3585 participants. Women were older, had a higher body mass index (BMI), and were shorter. The prevalence of lower limb disability and hip pain was twice as great in women as in men. Men showed a higher prevalence than women when OA was defined by Kellgren and Lawrence or Croft grades. Of the subjects with hip pain, 98.8% had had pain for more than a month; of these, 30.2% had pain for between one and five years and 51.1% for longer than five years. The prevalence of ROA was a much higher when defined by Croft grade 3 than by the other definitions

**Table 4** Association between the three different definitions of radiographic hip osteoarthritis studied and clinical symptoms, stratified for sex (n = 3561)

	Hip pain		Morning stiffness		LLD $\geq$ 0.5		LLD $\geq$ 1.0	
	Men	Women	Men	Women	Men	Women	Men	Women
<b>Kellgren</b>								
$\geq$ grade 2	1.6* (0.8 to 3.2)	3.5 (2.3 to 5.2)	1.1 (0.7 to 1.7)	1.3 (0.9 to 1.9)	1.5* (0.8 to 3.0)	3.3 (2.2 to 4.9)	1.0* (0.3 to 2.7)	4.6 (2.9 to 7.3)
$\geq$ grade 3	8.7 (3.1 to 24.5)	5.7 (2.7 to 12.0)	2.5 (0.9 to 6.9)	1.9 (0.9 to 4.1)	4.4 (1.4 to 13.9)	8.7 (4.1 to 18.6)	3.6 (1.8 to 16.4)	9.6 (4.4 to 20.7)
<b>MJS</b>								
$\leq$ 2.5 mm	2.7 (1.5 to 5.0)	2.3 (1.5 to 3.4)	1.6 (1.0 to 2.5)	1.5 (1.1 to 2.2)	2.5 (1.4 to 4.6)	2.7 (1.9 to 4.0)	1.2* (0.4 to 3.2)	3.8 (2.4 to 5.8)
$\leq$ 2.0 mm	4.4 (1.9 to 10.0)	4.4 (2.6 to 7.5)	1.6 (0.8 to 3.3)	1.7 (1.0 to 2.9)	2.2 (0.8 to 5.8)	4.4 (2.6 to 7.4)	1.6 (0.4 to 6.6)	5.1 (2.8 to 9.2)
$\leq$ 1.5 mm	6.6 (2.2 to 19.4)	6.6 (3.1 to 14.0)	1.9 (0.7 to 5.1)	2.1 (1.0 to 4.4)	3.0 (0.9 to 10.9)	6.5 (3.1 to 13.9)	3.7 (0.8 to 16.8)	7.1 (3.2 to 15.7)
<b>Croft grade</b>								
$\geq$ grade 3	1.3 (0.9 to 2.0)	1.5 (1.1 to 2.0)	0.8 (0.7 to 1.1)	0.9 (0.8 to 1.1)	0.8 (0.5 to 1.3)	1.2 (0.9 to 1.6)	0.8 (0.4 to 1.4)	1.4 (1.0 to 2.0)
$\geq$ grade 4	2.9 (1.4 to 6.0)	4.1 (2.5 to 6.5)	1.4 (0.8 to 2.5)	1.7 (1.1 to 2.6)	2.7 (1.3 to 5.5)	3.8 (2.4 to 6.1)	1.3* (0.4 to 4.4)	5.1 (3.0 to 8.6)

Associations are presented by odds ratios (95% confidence intervals in parentheses). LLD, lower limb disability index (a score of  $\geq$ 0.5 was defined as moderately disabled; a score of  $\geq$ 1.0 as severe disabled). Associations are adjusted for body mass index and radiographic osteoarthritis of the other hip. \*Significant difference between men and women.

**Table 5** Association between two definitions of radiographic hip osteoarthritis and hip pain, stratified for sex and age

	Men		Women		Total	
	Younger	Older	Younger	Older	Younger	Older
Median age (years)	60.8 (n=776)	70.1 (n=723)	60.5 (n=1016)	71.1 (n=1070)	60.6 (n=1792)	70.6 (n=1793)
<i>Kellgren</i>						
≥grade 2	0.3* (0.0 to 1.9)	3.7 (1.7 to 7.9)	4.0 (2.0 to 8.2)	3.2 (1.9 to 5.3)	1.7 (1.0 to 3.2)	3.2 (2.1 to 4.9)
Proportion with hip pain (%)	2.1	16.9	36.1	31.7	16.7	25.5
<i>MJS</i>						
≤2.5 mm	1.6 (0.5 to 4.7)	4.1 (1.9 to 8.8)	2.2 (1.1 to 4.2)	2.3 (1.4 to 3.8)	2.0 (1.2 to 3.5)	2.7 (1.8 to 4.2)
Proportion with hip pain (%)	10.8	17.9	23.6	25.5	18.5	22.7

Associations are presented as odds ratios (95% confidence intervals in parentheses). Men and women are divided in two equal groups, a younger and an older, divided by the median age (65.2 years). Associations are adjusted for body mass index and radiographic osteoarthritis of the other hip. \*48 cases with radiographic hip osteoarthritis according to Kellgren and Lawrence<sup>5</sup>; only one of these reported hip pain.

of moderate hip OA. The prevalence of moderate radiological hip OA defined by Kellgren and Lawrence and MJS was similar.

**Reliability**

Table 2 shows the inter-rater reliability for different individual radiological features and three definitions of hip ROA. The inter-rater reliability for the different individual radiological features was relatively low, with the exception of the MJS assessed as a continuous variable. Kellgren and Lawrence ≥grade 2 and MJS ≤2.5 mm had comparable reliability, whereas for Croft ≥grade 3 the reliability was somewhat lower.

**Construct validity**

Table 3 shows the association between the three definitions of hip ROA for different cut off points and clinical symptoms of hip OA—hip pain, morning stiffness, and lower limb disability (moderate and severe). The percentages of subjects defined in these ways according to the different cut off points are given in table 1. Table 3 shows that severe hip ROA had a stronger association with symptoms than moderate hip ROA. The Kellgren and Lawrence grade and MJS demonstrated comparable associations with clinical symptoms of hip OA, especially with hip pain and lower limb disability for both moderate and severe hip ROA. The Croft grade had the weakest associations with clinical symptoms of hip OA.

**Sex as an effect modifier**

We found that men had on average a larger joint space width than women (4.2 v 3.9 mm, respectively). Furthermore we

found that height was positively correlated with joint space width. We also found a positive correlation within sex between height and the joint space width (respectively, a β value of 0.16 for men and 0.14 for women). Because in the present study women were shorter than men, we adjusted for height. After adjustment for height, the sex effect disappeared.

Table 4 shows the association between different definitions of hip ROA and the clinical symptoms of hip OA stratified for sex. For both definitions of Croft, the results showed no significant sex difference except for the association between Croft ≥grade 4 and severe lower limb disability. We found, however, that sex was a significant effect modifier for Kellgren and Lawrence grade (≥grade 2). For women the associations between symptoms (hip pain and lower limb disability) and hip OA defined by Kellgren and Lawrence (≥grade 2) were significantly stronger than for men. We also found that for women the association between symptoms and hip ROA according to Kellgren and Lawrence (≥grade 2) was stronger than according to the MJS (≤2.5 mm) or Croft grade (≥grade 3). Women had a higher BMI than men, but after we adjusted for BMI the assessed associations did not change.

Table 5 shows the association between different definitions of hip ROA (Kellgren ≥grade 2 and MJS ≤2.5 mm) and hip pain stratified for sex and age (two categories). We divided men and women into two equal groups, a younger and older group (around the median of 65.2 years). Older persons had a stronger association between hip ROA and hip pain than younger persons, especially when defined by Kellgren and Lawrence. The trend was that hip ROA in younger men,

**Table 6** Predictive validity of the three definitions for total hip replacement (THR) at follow up (n=3561)

	Total incident THR at follow up*	THR/ROA†	OR (95% CI)
<i>Kellgren and Lawrence (≥grade 2)</i>			
Right	57	33/154 (21.4)	30.6 (17.5 to 53.5)
Left	42	22/110 (20.0)	34.3 (18.1 to 65.2)
<i>MJS (≤2.5 mm)</i>			
Right	57	26/151 (17.2)	18.6 (10.7 to 32.3)
Left	42	19/123 (15.4)	22.6 (11.8 to 43.0)
<i>Croft grade (≥grade 3)</i>			
Right	57	46/673 (6.8)	16.0 (8.0 to 31.8)
Left	42	28/717 (3.9)	6.7 (3.4 to 12.9)

\*Number of incident THR at follow up predicted by each definition/number ROA cases at baseline, as defined by each definition of hip ROA.

†Number of cases of THR predicted by each definition/number of ROA cases defined by each definition. CI, confidence interval; OR, odds ratio; ROA, radiological osteoarthritis.

especially when defined by Kellgren and Lawrence, had a weaker relation with hip pain than in women (both age categories) or older men. These results were similar for the association with lower limb disability.

### Predictive validity

Table 6 shows the predictive validity of the three definitions for THR at follow up, indicated by the association between the different definitions of hip ROA at baseline and THR at follow up. The Kellgren and Lawrence grading system predicted the highest ratio of number of incident THR at follow up divided by the number ROA cases at baseline, and showed the strongest association with THR at follow up, compared with the other definitions.

## DISCUSSION

Based on the results of the present study, with the Kellgren and Lawrence grade being the best predictor for a THR at follow up and MJS being height dependent, we concluded that radiological hip OA might be better defined for epidemiological studies by the Kellgren and Lawrence system than by MJS.

The inter-rater reliability of Kellgren and Lawrence grade assessed in this study is similar to that described in published reports.<sup>4 6 22 23</sup> In contrast to more recent studies, the original study of Kellgren and Lawrence showed a relatively low inter-rater reliability (ICC of 0.40).<sup>10</sup> In the present study we found an inter-rater reliability of the MJS according to Croft that was similar to previous studies.<sup>4 7 22 24 25</sup> The inter-rater reliability for Croft grade  $\geq 3$  in the present study had a  $\kappa$  value of 0.51 compared with values of 0.37 to 0.79 in earlier studies.<sup>4 7 22 26</sup> The wide range of inter-rater reliability between these studies is mainly explained by the different cut off levels used. One study<sup>14</sup> used the same cut off level as the present study, and reported a similar  $\kappa$  value of 0.41. However, in the original study by Croft the  $\kappa$  values presented were based on measurement of the size of the individual radiological features and not on atlas based grades. The inter-rater reliability reported in the present study was similar for subchondral sclerosis and for osteophytes compared with the reliability reported in the study by Croft.<sup>14</sup>

The validity of the different definitions of hip ROA has been poorly investigated in previous studies. In the present study we investigated the construct and predictive validity. Because of the absence of a gold standard we expressed the predictive validity as the ability of each definition to predict a THR at follow up. The requirement for a THR has been proposed as a potential outcome measure based on the assumption that THR is undertaken only in patients with severe disease—from both a symptomatic (painful and disabling disease) and a structural point of view (overall severity or advanced joint space narrowing).<sup>27 28</sup> The lower limb disability assessed by the health assessment questionnaire (HAQ) in the Rotterdam study is not a disease specific outcome measure, but it measures arthritic conditions in general. On the other hand, lower limb disability is an important symptom of hip OA, and OA is the most important cause of disability of elderly people.<sup>29</sup> Hence we included lower limb disability, assessed by the HAQ—as well as the presence of hip pain and morning stiffness—as an important symptom of hip OA in the analysis. Overall, the Kellgren and Lawrence grading system had the best predictive validity when compared with the other definitions of hip ROA, and similar associations with symptoms of hip OA (construct validity) as MJS. MJS came out better in both construct and predictive validity than Croft grade. The weak associations reported in the present study between Croft grade  $\geq 3$  and symptoms of hip OA can be explained by the high prevalence of moderate hip OA, and presumably therefore by the low

specificity value, using hip pain as the gold standard for Croft grade  $\geq 3$ . Therefore it is difficult to compare the Croft definition (moderate hip OA,  $\geq$  grade 3) with the other definitions. An earlier study reported similar prevalence of hip OA defined by Croft grade  $\geq 3$  and MJS  $\leq 2.5$  mm, and also a similar prevalence of hip pain in “disease positive” hips.<sup>14</sup> When we excluded those subjects with an incident hip fracture during the follow up period, and repeated the analysis for predictive validity, the results did not change essentially.

The second objective of our study was to investigate whether the relation between the three definitions and symptoms was sex dependent. Surprisingly, only the strength of the association between the Kellgren and Lawrence grading system and symptoms of hip OA was sex dependent. These findings have not been reported in previous studies. A possible explanation for this sex dependency could be the stronger relation between (femoral) osteophytes and hip pain in women. In women we found a stronger relation between osteophytes and hip pain (odds ratio of 1.7 for women *v* 1.2 for men); however, the prevalence of osteophytes in women was lower (34.3% in women *v* 43.6% in men). In contrast to our findings, we had expected that the strength of the association between MJS and symptoms of hip OA would have been sex dependent.

A stronger association between hip ROA and hip pain was found in older persons than in younger ones. Because of a power problem (owing to the smaller sample size for the younger male category) the sex difference was not significant. A possible explanation for this difference might be that younger persons have better muscle strength in the lower limb than older persons. Reduced muscle strength is regarded as a risk factor for pain and disability in OA,<sup>30–32</sup> and exercise therapy aiming to improve muscle strength has a beneficial effect on pain in patients with OA of the hip or knee.<sup>33 34</sup>

The results of the present study may be flawed by the quality of the radiographs. In particular, measurement of joint space width (MJS) could be flawed because of the quality of the radiographs. Important variables in the radiographic procedures are the position of the central ray of the x ray beam relative to the centre of the joint, and the distance between the centre of the joint and the x ray film (focus to film distance). Centring the x ray beam on the umbilicus instead of on the superior aspect of the symphysis pubis resulted in an average increase in joint space width of about 10%.<sup>16</sup> The focus to film distance may also modify the measurement.<sup>35</sup> On the other hand in the Croft's study<sup>14</sup> the x ray beam was also centred 10 cm higher than a standard anteroposterior view of the pelvis.

The source of potential bias in this study is a probable health based selection. The subjects in the present study had to be mobile enough to visit the research centre at baseline and follow up and to survive the follow up period (mean 6.6 years). Overall, participants were generally healthier than non-participants. In other words, patients with the most severe symptoms were most probably not included. It seems likely that, in this younger and healthier population with less frequent lower limb disability and hip pain, the prevalence of hip ROA is underestimated, as may be the magnitude of the association between the different definitions of hip ROA and symptoms of hip OA. Knowing that for older persons a stronger relation was found between hip ROA and hip pain, especially when defined by Kellgren and Lawrence, this underestimation may particularly hold for Kellgren and Lawrence grade.

When we compared the results of Kellgren and Lawrence grade and MJS we found the following differences. Kellgren and Lawrence was the best predictor for a THR at follow up. As described earlier by Buckland-Wright<sup>36</sup> and Lanyon *et al*,<sup>37</sup>

**Table 7** The Kellgren and Lawrence grading system

Grade	Description
0	No osteoarthritis
1	Doubtful Possible narrowing of joint space medially and possible osteophytes around the femoral head; or osteophytes alone
2	Mild Definite narrowing of joint space inferiorly, definite osteophytes, and slight sclerosis
3	Moderate Marked narrowing of joint space, definite osteophytes, some sclerosis and cyst formation, and deformity of the femoral head and acetabulum
4	Severe Gross loss of joint space with sclerosis and cysts, marked deformity of femoral head and acetabulum and large osteophytes

Hips classified as grade 2 or more were defined as having osteoarthritis.

**Table 8** Croft's modification of the Kellgren and Lawrence grading system ("Croft grade")

Grade	Description
0	No change
1	Definite osteophytes only
2	JSN only (defined as an MJS of $\leq 2.5$ mm)
3	Presence of two of the following: JSN, osteophytosis, subchondral sclerosis (of $\geq 5$ mm), cyst formation
4	Presence of three of the following: JSN, osteophytosis, subchondral sclerosis (of $\geq 5$ mm), cyst formation
5	Same as grade 4, but with deformity of the femoral head or total hip replacement due to OA (verified by record view)

JSN, joint space narrowing; MJS, minimal joint space; OA, osteoarthritis.

**Table 9** Croft's measurement of the minimal joint space (lateral, superior, axial, medial)

Grade	Description
0	MJS $> 2.5$ mm
1	MJS $> 1.5$ mm and $\leq 2.5$ mm
2	MJS $\leq 1.5$ mm

Minimal joint space is the shortest distance on the radiograph between the femoral head margin and the acetabular edge.

we also found that men had larger joint spaces than women. After adjustment for height these joint space differences between men and women disappeared. Considering these results, it is doubtful whether the given cut off point of MJS is valid for people of short stature.

When we stratified the associations between each definition and the symptoms of hip OA (hip pain and lower limb disability) for sex, surprisingly we found significantly stronger associations for Kellgren and Lawrence grade for women with hip pain and lower limb disability than for men.

Based on these results, we concluded that the Kellgren and Lawrence grade is still a useful definition of hip ROA for epidemiological studies.

## ACKNOWLEDGEMENTS

This study was supported by a grant from the Dutch Arthritis Association. We are very grateful to F van Rooij, E van der Heijden, R Vermeeren, and L Verwey for collection of follow up data. Moreover, we thank the participating general practitioners, the pharmacists, the many field workers at the research centre in Ommoord, and of course all the participants.

## Authors' affiliations

**M Reijman**, R M D Bernsen, B W Koes, S M A Bierma-Zeinstra, Department of General Practice, Erasmus MC, Rotterdam, Netherlands  
**H A P Pols**, Department of Internal Medicine, Erasmus MC, Rotterdam  
**J M W Hazes**, Department of Rheumatology, Erasmus MC, Rotterdam

## APPENDIX

The Kellgren and Lawrence grading system is shown in table 7, Croft's modification of the Kellgren and Lawrence grading system in table 8, and Croft's measurement of the minimal joint space in table 9.

## REFERENCES

- 1 **Felson DT**, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 1998;**41**:1343-55.
- 2 **Birrell F**, Croft P, Cooper C, Hosie G, Macfarlane G, Silman A. Health impact of pain in the hip region with and without radiographic evidence of osteoarthritis: a study of new attenders to primary care. *Ann Rheum Dis* 2000;**59**:857-63.
- 3 **Sun Y**, Gunther K, Brenner H. Reliability of radiographic grading of osteoarthritis of the hip and knee. *Scand J Rheumatol* 1997;**26**:155-65.
- 4 **Hirsch R**, Fernandes RJ, Pillemer SR, Hochberg MC, Lane NE, Altman RD, et al. Hip osteoarthritis prevalence estimates by three radiographic scoring systems. *Arthritis Rheum* 1998;**41**:361-8.
- 5 **Gunther KP**, Sun Y. Reliability of radiographic assessment in hip and knee osteoarthritis. *Osteoarthritis Cartilage* 1999;**7**:239-46.
- 6 **Reijman M**, Hazes JMW, Koes BW, Verhagen AP, Bierma-Zeinstra SMA. Validity, reliability and applicability of seven definitions of hip osteoarthritis used in epidemiological studies: a systematic appraisal. *Ann Rheum Dis* 2004;**63**:226-32.
- 7 **Croft P**, Cooper C, Coggon D. Case definition of hip osteoarthritis in epidemiologic studies. *J Rheumatol* 1994;**21**:591-2.
- 8 **Hofman A**, Grobbee D, de Jong PT, VandenOuweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991;**7**:403-22.
- 9 **Oidding E**, Valkenburg H, Algra D, VandenOuweland F, Grobbee D, Hofman A. Associations of radiological osteoarthritis of the hip and knee with locomotor disability in the Rotterdam Study. *Ann Rheum Dis* 1998;**57**:203-8.
- 10 **Kellgren JH**, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;**16**:494-502.
- 11 **Kellgren JH**, Jeffrey MR, Ball J. *The epidemiology of chronic rheumatism. Atlas of standard radiographs of arthritis*. Oxford: Blackwell Scientific Publications, 1963.
- 12 **Altman RD**, Hochberg MC, Murphy WA, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995;**3**:3-70.
- 13 **Burnett S**, Hart DJ, Cooper C, Spector TD. *A radiographic atlas of osteoarthritis*. London: Springer-Verlag, 1994.
- 14 **Croft P**, Cooper C, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol* 1990;**132**:514-22.
- 15 **Auleley G**, Rousselin B, Ayral X, Edouard-Noel R, Dougados M, Ravaud P. Osteoarthritis of the hip: agreement between joint space width measurements on standing and supine conventional radiographs. *Ann Rheum Dis* 1998;**57**:519-23.

- 16 **Auleley GR**, Duche A, Drape JL, Dougados M, Ravaud P. Measurement of joint space width in hip osteoarthritis: influence of joint positioning and radiographic procedure. *Rheumatology* 2001;**40**:414–19.
- 17 **Ravaud P**, Chastang C, Auleley GR, Giraudeau B, Royant V, Amor B, *et al*. Assessment of joint space width in patients with osteoarthritis of the knee: a comparison of 4 measuring instruments. *J Rheumatol* 1996;**23**:1749–55.
- 18 **Donner A**, Eliasziw M. A goodness-of-fit approach to inference procedures for the Kappa statistic: confidence interval construction, significance testing and sample size estimation. *Stat Med* 1992;**11**:1511–19.
- 19 **Walter SD**, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;**17**:101–10.
- 20 **Buchbinder R**, Goel V, Bombardier C, Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *J Clin Epidemiol* 1996;**49**:141–9.
- 21 **Felson DT**, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. *Baillieres Clin Rheumatol* 1995;**9**:253–66.
- 22 **Ingvarsson T**, Hagglund G, Lindberg H, Lohmander L. Assessment of primary hip osteoarthritis: comparison of radiographic methods using colon radiographs. *Ann Rheum Dis* 2000;**59**:650–3.
- 23 **Bierma-Zeinstra SMA**, Bohnen AM, Ginai AZ, Prins A, Verhaar JAN. Validity of American College of Rheumatology criteria for diagnosing hip osteoarthritis in primary care research. *J Rheumatol* 1999;**26**:1129–33.
- 24 **Antoniades L**, MacGregor AJ, Matson M, Spector TD. A cotwin control study of the relationship between hip osteoarthritis and bone mineral density. *Arthritis Rheum* 2000;**43**:1450–5.
- 25 **Birrell F**, Croft P, Cooper C, Hosie G, Macfarlane G, Silman A. Radiographic change is common in new presenters in primary care with hip pain. *Rheumatology* 2000;**39**:772–5.
- 26 **MacGregor A**, Antoniadis L, Matson M, Andrew T, Spector TD. The genetic contribution to radiographic hip osteoarthritis in women: Results of a classic twin study. *Arthritis Rheum* 2000;**43**:2410–16.
- 27 **Fox KM**, Hochberg MC, Resnik CS, Kenzora JE, Hebel JR, Zimmerman SI, *et al*. Severity of radiographic findings in hip osteoarthritis associated with total hip arthroplasty. *J Rheumatol* 1996;**23**:693–7.
- 28 **Dougados M**, Guegen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? *J Rheumatol* 1999;**26**:855–61.
- 29 **Brazier JE**, Harper R, Munro J, Walter SJ, Snaithe ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;**38**:870–7.
- 30 **van Baar ME**, Dekker J, Lemmens JA, Oostendorp RA, Bijlsma JW. Pain and disability in patients with osteoarthritis of hip or knee: the relationship with articular, kinesiological, and psychological characteristics. *J Rheumatol* 1998;**25**:125–33.
- 31 **Dekker J**, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. *J Behav Med* 1992;**15**:189–214.
- 32 **Stuifjens MP**, Dekker J, van Baar ME, Oostendorp RA, Bijlsma JW. Muscle strength, pain and disability in patients with osteoarthritis. *Clin Rehabil* 2001;**15**:331–41.
- 33 **van Baar ME**, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. *Arthritis Rheum* 1999;**42**:1361–9.
- 34 **van Baar ME**, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. *Ann Rheum Dis* 2001;**60**:1123–30.
- 35 **Ravaud P**, Dougados M. Radiographic assessment in osteoarthritis. *J Rheumatol* 1997;**24**:786–91.
- 36 **Buckland-Wright J**. Quantitative radiography of osteoarthritis. *Ann Rheum Dis* 1994;**53**:268–75.
- 37 **Lanyon P**, Muir K, Doherty S, Doherty M. Age and sex differences in hip joint spaces among asymptomatic subjects without structural change. *Arthritis Rheumatism* 2003;**48**:1041–6.