

## REVIEW

# Validity, reliability, and applicability of seven definitions of hip osteoarthritis used in epidemiological studies: a systematic appraisal

M Reijman, J M W Hazes, B W Koes, A P Verhagen, S M A Bierma-Zeinstra

*Ann Rheum Dis* 2004;63:226–232. doi: 10.1136/ard.2003.010348



The appendixes can be seen on the website at <http://www.annrheumdis.com/supplemental>

See end of article for authors' affiliations

Correspondence to: M Reijman, MSc, Department of General Practice, Erasmus MC-Faculty, PO Box 1738, 3000 DR Rotterdam, The Netherlands; [m.reijman@erasmusmc.nl](mailto:m.reijman@erasmusmc.nl)

Accepted 19 May 2003

**Objective:** To summarise and review articles addressing the quality (validity, reliability, applicability) of seven commonly used definitions of hip osteoarthritis (OA) for epidemiological studies in order to use it primarily as a classification criterion.

**Methods:** Medline and Embase were searched and articles studying the validity, reliability, or applicability of the definitions of hip OA were selected. Two reviewers independently extracted data on the quality of the seven definitions.

**Results:** Review of the literature showed the validity of the various definitions of hip OA, in particular, has barely been investigated. Minimal joint space (MJS) demonstrated the highest (intra- and interrater) reliability, and showed the highest association with hip pain and restricted internal rotation compared with the other definitions of hip OA. The reliability of the Kellgren and Lawrence grade and the index according to Lane is comparable with that of the MJS, but the construct validity should be investigated more thoroughly. The reliability and validity according to the Croft grade were inferior to the MJS, the Kellgren and Lawrence grade, and the index according to Lane. Despite precise and extensive development, the ACR criteria showed poor reliability and poor cross-validity (agreement between three ACR criteria sets) in a primary care setting.

**Conclusions:** The reliabilities of MJS, Kellgren and Lawrence, and the index according to Lane were comparable, but the MJS had the highest relationship with hip pain in a male population. Considering how often definitions of hip OA are used, it is surprising that the validity has been so poorly investigated, and the validity needs to be studied more thoroughly.

Osteoarthritis (OA) is the most common joint disorder<sup>1</sup> and is a considerable burden on society. Depending on the definition of hip OA used, the prevalence ranges from 7 to 25% in people aged 55 years and over.<sup>2</sup> The hip is particularly interesting because it is often the sole joint affected by OA, suggesting that local biomechanical risk factors are important. In addition, the hip is crucial to independent function.<sup>3</sup>

A major problem in studying hip OA, is the absence of consensus in defining hip OA for epidemiological and clinical studies.<sup>4</sup> Most epidemiological studies have used a single hallmark of hip OA—namely, radiological changes, to define hip OA.<sup>5,6</sup>

To investigate (potential) risk factors a valid and reliable definition of hip OA is required. Therefore we appraised the quality—that is, the validity, reliability, and applicability, of seven definitions of hip OA commonly used for epidemiological studies:

1. The radiological grading system of Kellgren and Lawrence<sup>7</sup>
2. Croft's radiological grading system (a modification of the Kellgren and Lawrence grading system)<sup>8</sup>
3. Minimal joint space (MJS) according to Croft *et al* (a measurement of the narrowing of the joint space)<sup>8</sup>
4. Measurement of the joint space according to Resnick and Niwayama<sup>9</sup>
5. Three sets of criteria (one clinical, and two combined sets of clinical and radiographic criteria) of the American College of Rheumatology (ACR)<sup>10</sup>

6. Clinical definition of hip OA: radiological OA combined with pain in the hip region<sup>11,12</sup>
7. Radiographic index grade according to Lane.<sup>13,14</sup>

The objective of our study was to review the quality (reliability, validity, applicability) of these seven definitions of hip OA commonly used in epidemiological studies, in order to use it primarily as a classification criterion.<sup>15,16</sup>

## METHODS

The literature was searched for all relevant papers containing one of the seven definitions of hip OA. Studies which fulfilled predefined inclusion criteria were identified and subsequently assessed for aspects of reliability, validity, and applicability of the definition of hip OA used in each particular study.

## Identification of the literature

To identify the studies a search was made in the following databases: Medline/Pubmed (1966–March 2002), Cochrane Library and Embase (1990–March, 2002). The specific keywords were: “osteoarthritis, hip” or “osteoarthritis” and “hip” and “clinical definition”, “radiological definition”, “case definition”, “radiographic grading”, “diagnosis”, “severity”, “index of severity”, “classification criteria”,

**Abbreviations:** ACR, American College of Rheumatology; ICC, intraclass correlation coefficient; MJS, minimal joint space; OA, osteoarthritis; ROM, range of movement

**Table 1** Characteristics of the included studies (n = 14)

Reference	Definition used in study	Study group				Qualification of observers	Radiograph	Prevalence (%)	Number of observers	Number of readings
		Setting	Size	Male (%)	Age					
Allman (1991) <sup>10</sup> Antoniades (2000) <sup>22</sup> (Croff)	ACR St Thomas' UK twin register	227	63.7	61 (mean)	Musculoskeletal radiologist Trained	AP Pelvic AP (supine)	NA 8 (≥2) 5 (≥3) 8 (≤2.5 mm)	1	2	
		1148	0	53 (median) (range 40-70)						
Bellamy (1999) <sup>23</sup> Bierma-Zeinstra (1999) <sup>12</sup>	ACR Australian twin register General practice	159	28.9	58.79 (mean)	Rheumatologist Radiologist and 1 trained	AP Pelvic AP and axial (frog-leg position)	NP NA*	3	1	
		227	27	66 (mean) (SD 9.6) (range 40-70)						
Birrell (2001) <sup>20</sup> Croff (1990) <sup>8</sup> (Croff)	ACR General practice Intravenous urogram	195	32.3	63 (median) (IQR 54-71)	NP Trained	Pelvic AP Intravenous urogram	NA* 24.3 (≥2) 11.0 (≥3) 14.4 (≤2.5) 2.0 (≤1.5) 3.6 10.8 (<2.5)	2	2	
		1315	100	60-75 (range)						
Hirsch (1998) <sup>5</sup> Ingvarsson (2000) <sup>19</sup> (Croff)	Kellgren and Lawrence MUS (Croff) Population based Colon radiographs	755	37.9	58.3 (mean)	Trained NP	Pelvic AP Colon AP (supine)	9.2 10.0 (≤2.5)	3	2	
		1501	42.5	Not given						
Kellgren (1957) <sup>7</sup> Lane (1993) <sup>13</sup> Lane	Kellgren and Lawrence Index according to fractures St Thomas UK adult twin register	85	NP	55-64 (range) 70 (median)	Trained Trained	NP Pelvic AP	NP 55.5 (right hip ≥2) 36.1 (left hip ≥2) 11.7 (≥2) 10.9	2	2	
		31	0	(range 65-86) 57.4 (mean) (range 50-72)						
MacGregor (2000) <sup>21</sup> Nevitt (1995) <sup>14</sup> Lane	ACR Study of osteoporotic fractures Intravenous urogram	824	0	57.4 (mean) (range 50-72)	Trained Trained	Pelvic AP (supine) Pelvic AP	11.7 (≥2) 10.9 7.1 (≥2) 4.7 (≥3)	1	2	
		4090	0	71.1 (mean) (SD 4.9) 60-75 (range)						
Smith (1995) <sup>27</sup> Sun (1997) <sup>4</sup> (Croff)	Kellgren and Lawrence MUS (Croff) Systematic review	203	0	60-75 (range)	NP -	Intravenous urogram -	NP -	2	2	
		-	-	-						

\*Study group includes patients with hip pain who consulted the general practitioner. NA, not applicable; NP, not provided; AP, anteroposterior radiograph; IQR interquartile range.

**Table 2** Reliability of the definitions of hip osteoarthritis

	Intrater	Interrater	Cut off level	Statistic	Size	Prevalence(%)	Reference
Kellgren and Lawrence (0-4)	0.66-0.89	0.75	≥2	κ (dichotomous)	227	NA	Bierma-Zeinstra (1999) <sup>12</sup> Hirsch (1998) <sup>5</sup>
		0.63	-	Intra: Pearson	775	3.6	
	0.76 0.75	0.60-0.65	≥2	κ (dichotomous)	147	9.2	Ingvarsson (2000) <sup>19</sup> Kellgren (1957) <sup>7</sup>
		0.40	-	Correlation coefficient	85	NP	
Croft Grade (0-5)	0.81 0.49-0.93	0.79	NP	κ		NA	Birrell (2001) <sup>20</sup> Croft (1990) <sup>8</sup>
		0.41-0.63	≥3 ≥4	κ (dichotomous)	50	24.3 (≥2) 11.0 (≥3)	
	0.61 0.81	0.37	NP	κ	350	16.3 (≥1) 11.7 (≥2)	Hirsch (1998) <sup>5</sup> MacGregor (2000) <sup>21</sup>
			NP	κ			
	0.61	0.37	≥3	κ	40	NP	Smith (1995) <sup>27</sup>
MJS according to Croft		0.81-0.84	NP	κ	350	8*	Antoniades (2000) <sup>22</sup> Birrell (2001) <sup>20</sup>
	0.83	0.75	-	ICC	195	NA	
	0.81	0.70	≤ 2.5 mm	κ (dichotome)	50	14.4	Croft (1990) <sup>8</sup>
	0.83	0.79	≤ 1.5 mm	κ		2.0	
	0.85	0.42	NP	κ			Hirsch (1998) <sup>5</sup>
	0.94	0.81/0.96	-	ICC	147	10	Ingvarsson (2000) <sup>19</sup>
	0.84		NP	κ	350	10.9	MacGregor (2000) <sup>21</sup>
0.85	0.42	≤ 2.5 mm	κ	40	NP	Smith (1995) <sup>27</sup>	
ACR criteria (dichotomous)							
	Clinical	0-0.65 (0.93-0.95†)		κ (dichotomous)	159	NP	Bellamy (1999) <sup>23</sup>
Clin/rad /lab	0.31-0.85 (0.91-0.97†)			κ (dichotomous)	159	NP	Bellamy (1999) <sup>23</sup>
Index according to Lane	0.70/0.88	0.76/0.85	0-3	ICC	31	55.5 (right hip ≥2) 36.1 (left hip ≥2)	Lane (1993) <sup>13</sup>
	0.36	0.40/0.47	≥1	κ			
	0.83	0.72/0.92	≥2	κ			
		0.87	0-4	ICC	40	7.1 (≥2)	Nevitt (1995) <sup>14</sup>
		0.71	≥2	κ	90	4.7 (≥3)	

\* <2.5 mm used as cut off level; † κ adjusted for prevalence and bias.<sup>24</sup>

NP, not provided; NA, not applicable (group includes patients with hip pain who consulted the general practitioner); ICC, intraclass correlation coefficient.

“radiographic change”, “minimal joint space”, “Kellgren”, “Kellgren and Lawrence”, “reliability”, “reproducibility of results”, “epidemiologic studies”, or “feasibility studies”. The search was extended by screening the reference lists of all relevant articles identified. We repeated the search using the keywords of all selected articles.

### Criteria for studies considered for inclusion

A study was included in this review if it fulfilled all of the following criteria: (a) the study group contained people with and without hip OA; (b) it was an original article or a systematic review; (c) at least one of the seven definitions of hip OA investigated here was used; (d) the study described the design, or the reliability, or the validity, or the applicability of at least one of the above mentioned

definitions or the study investigated the risk factors (or determinants) of hip OA, and used at least two of the above mentioned definitions.

### Critical assessment of OA definitions

Using the information from the criteria of Buchbinder *et al*,<sup>15</sup> Felson *et al*,<sup>16</sup> and Bierma-Zeinstra *et al*,<sup>17</sup> we compiled a list of criteria to evaluate the definitions of hip OA (appendix 1, available on the website at <http://www.annrheumdis.com/supplemental>). These criteria relate to the reliability, validity, and applicability of the definition of hip OA:

1. The reliability of the definition expressed in intra- and interrater reliability
2. The validity of the definition expressed in

**Table 3** Association of definitions with known symptoms of hip osteoarthritis

	MJS (Croft)		Croft grade		
	≤ 2.5 mm	≤ 1.5 mm	≥ 2	≥ 3	≥ 4
Restriction ROM, OR (95% CI)					
Flexion (≤ 94°)		2.6 (0.8 to 8.9)	1.5 (0.7 to 3.2)		
External rotation (≤ 23°)		1.2 (0.3 to 3.9)	3.0 (1.4 to 6.2)		
Internal rotation (≤ 23°)		46.8 (5.2 to 420.0)	3.6 (1.6 to 8.0)		
Prevalence of pain (%)	28.3	56.0		28.8	47.5

References: 8, 20  
OR, odds ratio; CI, confidence interval.

### Criterion validity

- Expert validity: the expert validity evaluates the sensitivity and specificity of the classification criteria with the use of a predefined “gold standard” by expert’s opinion in a trans-sectional study design.<sup>15 16</sup>
- Predictive validity: the predictive validity evaluates the sensitivity and specificity of the classification criteria with the use of a predefined “gold standard” by an “obvious hip OA” (for example, a total hip replacement) after a certain period of follow up.<sup>15 16</sup>

### Construct validity

The construct validity evaluates whether the definition correlates with the external variables with which it should correlate.<sup>16 17</sup> In the case of radiological hip OA, the definition should correlate with known symptoms (hip pain, disability, limited range of movement (ROM), morning stiffness <1 hour) of hip OA, or with known risk factors of hip OA. If the definition is based on clinical signs, it should correlate with radiological signs of hip OA.

- The applicability of the definition of hip OA expressed in three issues—namely:
  - The ability to discriminate between hip OA and no hip OA
  - The ability to categorise the severity of hip OA
  - The tools and skills needed to define people with hip OA.
- A description of which method has been used to develop the definition of hip OA (content validity).

Two reviewers (MR and SMABZ) independently evaluated the definition of hip OA used in the included articles according to the above criteria. In cases of disagreement, both reviewers tried to achieve consensus. If the disagreement was not resolved, a third reviewer (BWK) was consulted to achieve a final judgment.

### Data extraction

In the included studies, data on reliability (various measures of intra- and interrater reliability), construct validity (association measures), and information on the applicability of the seven definitions used for hip OA were collected by two reviewers independently of each other and summarised (descriptive analysis) according to each definition separately.

## RESULTS

### Identification/selection of the literature

The initial searches resulted in 1170 potentially relevant articles.<sup>18</sup> Of these, 12 articles fulfilled the predefined inclusion criteria. After screening the reference lists of the 12 articles, another two articles were included. Finally, 14 publications were used to extract data on the reliability, validity, or applicability of the definitions for hip OA.

### Description of included studies

Of the 14 articles, 13 studied the reliability and 7 the validity of one (or more) of the definitions. Table 1 lists the characteristics of the studies. As can be seen, there is a large difference in the reported prevalence of hip OA, probably due to the large difference in the percentage of men and the different classifications of hip OA used. All studies used a relatively young population (mean age <66 years).

The 14 studies defined hip OA according to one (or more) of the following seven definitions (appendix 2; available on the website at <http://www.annrheumdis.com/supplemental>): Kellgren and Lawrence grade (five studies), Croft grade (six), MJS (according to Croft *et al*, eight.) MJS (according to Resnick and Niwayama, none), the ACR criteria (three), hip pain and joint space narrowing (one), and the index grade according to Lane (two).

## RESULTS OF THE INCLUDED STUDIES

### Reliability

Of the 14 studies, 13 investigated the reliability of five of the seven definitions of hip OA (table 2).

The four studies that investigated the reliability of the Kellgren and Lawrence grade reported an intrarater reliability with  $\kappa$  statistics of 0.76,<sup>19</sup> Pearson correlation coefficient of 0.66–0.89,<sup>5</sup> an interrater reliability with  $\kappa$  statistics of 0.60–0.75,<sup>12–19</sup> and an intraclass correlation coefficient (ICC) of 0.63.<sup>5</sup> In contrast with more recent studies, the original study of Kellgren and Lawrence showed a relatively lower interrater reliability (correlation coefficient of 0.40).<sup>7</sup>

In five studies the overall grade of Croft (a modification of the Kellgren and Lawrence grade) had an intrarater reliability with  $\kappa$  statistics of 0.49–0.93<sup>5 8 20 21</sup>, but a relatively lower interrater reliability with  $\kappa$  statistics of 0.37–0.79.<sup>5 8 20</sup> The wide range of intra- and interrater reliability between the studies is mainly explained by the different cut off levels used.

In seven studies the MJS according to Croft *et al* showed the highest intra- and interrater reliability compared with the other definitions of hip OA. The MJS according to Croft *et al* showed an intrarater reliability with  $\kappa$  statistics of 0.81–0.85<sup>5 8 21</sup> and an ICC of 0.83–0.94,<sup>19 20</sup> an interrater reliability with  $\kappa$  statistics of 0.42–0.84,<sup>5 8 22</sup> and an ICC of 0.75–0.96.<sup>19 20</sup>

**Table 4** Applicability of definitions of hip OA

	Kellgren	Croft grade	MJS (Croft)	ACR criteria	Hip pain+JSN	Index according to Lane
Discriminate +/-	Yes	Yes	Yes	Yes	Yes	Yes
Categorise severity	Yes (score: 0–4)	Yes (score: 0–5)	Yes (no/moderate/severe)	No	No	Yes (score 0–4)
Tests needed to perform definition	Radiographs	Radiographs	Radiographs	Clinical history/physical examination/radiographs/lab	Clinical history/Radiographs	Radiographs
Easy to perform (for MD level)	Yes	Yes	Yes	Yes	Yes	Yes

Only the study by Hirsch *et al*<sup>5</sup> described a relatively low interrater reliability with  $\kappa$  statistic of 0.42.

Only one study investigated the interrater reliability of the ACR classification(s)<sup>23</sup> and reported a wide range for the clinical set with  $\kappa$  statistics of 0.0–0.65 and the combined clinical, radiological, and laboratory signs set with  $\kappa$  statistics of 0.31–0.85.

Two studies investigated the reliability of the index according to Lane. These studies reported an intrarater reliability with  $\kappa$  statistics of 0.83 ( $\geq$  grade 2) and an ICC of 0.70–0.88,<sup>13</sup> an interrater reliability with  $\kappa$  statistics of 0.72–0.92 ( $\geq$  grade 2) and an ICC of 0.76–0.87.<sup>13, 14</sup>

### Validity

None of the screened studies investigated the criterion (expert or predictive) validity of the seven definitions of hip OA. In the 14 studies the construct validity was evaluated by considering two questions: Does the radiological definition correlate with known symptoms of hip OA? and Does the definition correlate with other definitions of hip OA? Of the 14 studies, seven evaluated the construct validity of two of the definitions of hip OA (MJS and the overall grade of Croft) (table 3). The association between the radiological definition and (known) symptoms of hip OA (hip pain, restricted ROM) was used as a measure of construct validity. The highest association was described between severe radiological hip OA and hip pain, and between severe radiological hip OA and a restricted internal rotation of the hip.<sup>8, 20</sup> In their study, Birrell *et al* investigated the association between restricted ROM and mild to moderate radiological hip OA defined as grade  $\geq 2$  (Croft grade) and severe OA defined as MJS  $\leq 1.5$  mm.<sup>20</sup> Internal rotation appeared to be the most discriminating movement for severe hip OA (OR = 46.8 (95% CI 5.2 to 420.0) *v* 3.6 (95% CI 1.6 to 8.0) for moderate OA). In 1990 Croft *et al* investigated the association between hip pain and radiological hip OA.<sup>8</sup> Severe hip OA defined by MJS  $\leq 1.5$  mm, showed a stronger association with hip pain than defined by the Croft grade (prevalence of 56.0% *v* 47.5% of those with hip pain). The association with pain and MJS  $\leq 2.5$  or Croft grade  $\geq 3$  is comparable (prevalence of 28.3% *v* 28.8% of those with hip pain).

For the construct validity we also reported the correlation between the different definitions of hip OA. The relationship between the Kellgren and Lawrence definition and the three sets of ACR criteria was very low ( $\kappa = 0.03$ –0.16).<sup>12</sup> There was a moderate agreement between the definition of Kellgren and Lawrence and “hip pain and joint space narrowing ( $\kappa = 0.52$ ).<sup>12</sup> There was a high association between a severe hip OA defined by MJS  $\leq 1.5$  mm and grade  $\geq 4$  (Croft grade) (OR = 153.5).<sup>8</sup> None of the studies compared the association between two of more definitions with known risk factors.

The method of development of the seven definitions of hip OA also differs considerably. The Kellgren and Lawrence grade and the index according to Lane were developed based on the opinion of the researchers. The overall grade of Croft and the MJS were based on a study group, and were developed according to pain within the study group. The ACR criteria sets were also based on a study group, and were developed using regression analysis (classification tree) on the occurrence of hip OA defined by an expert team. The methods of development of the remaining two definitions were not given.

### Applicability

The applicability of the definitions of hip OA in the present study was made operational as the ability to discriminate between hip OA and no hip OA, the ability to categorise the severity of hip OA, and the skills and tools needed to classify

people according to the respective definitions (table 4). According to their own description, six definitions are used to discriminate between people with and without hip OA, and all six are easy to apply for people at MD level. The Kellgren and Lawrence grade, Croft grade, the MJS and the index grade according to Lane can also categorise the severity of hip OA.

All definitions include information from a radiograph (except the clinical set of the ACR criteria). The ACR also makes use of information of the clinical history and physical examination (restricted ROM).

### DISCUSSION

Reviewing the selected literature demonstrates that the validity of the various definitions of hip OA, in particular, has barely been investigated. The highest (intra- and interrater) reliability was reported for the MJS and the index according to Lane, and the highest association with hip pain compared with the other definitions of hip OA for the MJS.

Despite putting much effort into identifying all relevant articles, some might have been missed because, for example, they used other keywords, had unclear abstracts, or were not indexed in Pubmed or Embase. Although the sensitivity of our search action might not be optimal,<sup>24, 25</sup> we nevertheless believe that we included the most appropriate studies that evaluated aspects of the quality of definition of hip OA, and assume that the data presented here give a clear insight into the currently available studies on this topic. Only 14 of 1170 potentially relevant articles fulfilled the predefined inclusion criteria. The most restrictive inclusion criterion was that the study group contained people with and people without hip OA.

The problems encountered when comparing the results of the included studies, were the differences in study groups (percentage of men), settings (open population, patients with hip pain who consulted their general practitioner), different cut off points for case definitions, and the different or non-transparent statistics used in the studies. For example, the percentage of men in the different studies ranged from 0 to 100%; because sex is a known risk factor for hip OA this will obviously influence the prevalence of hip OA. The prevalence, in turn, will also affect the value of reliability.<sup>26</sup> One study<sup>23</sup> adjusted the  $\kappa$  value (Cohen) they found for prevalence (prevalence adjusted bias adjusted  $\kappa$  /PABAK<sup>26</sup>); the adjusted  $\kappa$  was much higher than the crude  $\kappa$ .

In the absence of a “gold standard” for a definition of hip OA, we were particularly careful when evaluating the validity. Two potential solutions to define a “gold standard”, by expert’s opinion or by an “obvious hip OA” (such as total hip replacement) after a certain period of follow up were not used in the screened studies. Summarising the available information, it was clear that very few studies investigated the construct validity of the definitions used for hip OA. Of the 14 articles, not one focused on the relationship between risk factors and radiological hip OA, leaving us to evaluate the studies that reported the association between symptoms and radiological hip OA. Croft *et al* investigated the association between hip pain and radiological hip OA (two definitions of Croft)<sup>8</sup>; in their study group of 1315 men, only 759 completed the questionnaire (243 men died, 152 men were too ill according to the general practitioner). The men excluded were probably older, more disabled, and had more comorbidity than the men included, which might have led to a selection bias; the results of that study should therefore be interpreted with caution. Croft *et al* also investigated the association between individual radiological features and hip pain<sup>8</sup>; they concluded that MJS ( $\leq 1.5$  mm) showed a stronger association with hip pain than osteophytes (56% *v* 34.4%). Surprisingly, no articles were found that investigated

the association between the overall Kellgren and Lawrence grade and hip pain. The validity of three sets of criteria of the ACR was investigated in only one study,<sup>12</sup> which concluded that the clinical ACR criteria showed no cross-validity (agreement between three ACR criteria sets) with the two other ACR criteria sets, tested in primary care.

For reliability, the lack of comparability between the different studies is also an important confounder. Different standardisation of the x ray measurements between studies, or a possible difference in MJS between men and women, can influence the results of the reliability. Only one study directly compared the reliability of the Kellgren and Lawrence grade with MJS (according to Croft)<sup>19</sup>; the MJS showed a better (intra- and interrater) reliability. Five studies directly compared the overall grade of Croft and the MJS<sup>5, 8, 20, 21, 27</sup>, all these studies showed a better reliability of the MJS. No studies compared the other definitions. Only three studies reported the time interval between repeated readings: Croft *et al* 3–5 months<sup>8</sup>, Kellgren and Lawrence 1 month,<sup>7</sup> and Lane *et al* 1 month.<sup>13</sup> The length of this interval will probably influence the reliability (a longer time interval between repeat readings will give a lower intrarater reliability).<sup>4</sup> Overall, we assume that the MJS and the index according to the Lane definition for hip OA have the highest reliability for epidemiological and clinical studies.

The most commonly used definition of hip OA, the Kellgren and Lawrence grade, is also the one most criticised. Previous criticisms of the Kellgren and Lawrence grade include: inconsistencies in the description of radiographic features of OA,<sup>28–30</sup> the prominence awarded to osteophytes at all joint sites,<sup>1–30</sup> and a poor interrater and between-centre reliability.<sup>1, 28–30</sup> According to the articles included in our review, the interrater reliability was poor only in the original study of Kellgren and Lawrence,<sup>7</sup> but much better in three other much larger studies.<sup>5, 12, 19</sup> Notably, the same description of the Kellgren and Lawrence grade was used in all studies. Therefore in the present study we could not confirm the criticism of inconsistent grades and poor reliability of the Kellgren and Lawrence grade. The main criticism of the Kellgren and Lawrence grade is the importance of the presence of osteophytes. Although it is well known that the association between osteophytes and hip pain is poor,<sup>8</sup> not one of the 14 articles investigated the association between the overall Kellgren and Lawrence grade and hip pain. Overall, we assume that the Kellgren and Lawrence grade for hip OA is a useful definition for epidemiological studies.

Summarising the properties of the definitions used for hip OA investigated in the present study, we conclude that:

1. The MJS showed a good intra- and interrater reliability, a good association with hip pain and restricted internal rotation, and a good applicability; however, the quality (validity, reliability) of this definition should be investigated in an open population.
2. The Kellgren and Lawrence grade has a reliability comparable to that of the MJS, but the construct validity should be investigated more thoroughly.
3. The Croft grade appeared to be inferior to the MJS and the Kellgren and Lawrence grade for both reliability and validity.
4. The ACR criteria (despite their precise and extensive method of development) showed a poor reliability and a poor cross-validity in a primary care setting. Because these data are based on the results of only two studies, more research is needed on the ACR criteria (also in other settings).
5. The index according to Lane showed also a good intra- and interrater reliability, but no studies were included

which investigated the construct validity of this index grading system.

Considering how often the definitions of hip OA are used, it is surprising that the validity has been so poorly investigated. Meanwhile, because of the lack of such validity studies, we recommend that only those definitions with the best construct validity and the best reliability be used in epidemiological studies. We also recommend that the validity, especially the criterion (expert or predictive) validity, of the commonly used definitions be studied more thoroughly.

## ACKNOWLEDGEMENTS

This study was supported by a grant from the Dutch Arthritis Association.

## Authors' affiliations

**M Reijman, B W Koes, A P Verhagen, S M A Bierma-Zeinstra,** Department of General Practice, Erasmus MC, Rotterdam, The Netherlands

**M Reijman, J M W Hazes,** Department of Epidemiology and Biostatistics, Erasmus MC, Rotterdam, The Netherlands

**J M W Hazes,** Department of Rheumatology, Erasmus MC, Rotterdam, The Netherlands

## REFERENCES

1. **Croft P,** Cooper C, Coggon D. Case definition of hip osteoarthritis in epidemiologic studies. *J Rheumatol* 1994;**21**:591–2.
2. **Dougados M,** Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996;**55**:356–62.
3. **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 (Oxford)* 2000;**39**:772–5.
4. **Sun Y,** Gunther KP, Brenner H. Reliability of radiographic grading of osteoarthritis of the hip and knee. *Scand J Rheumatol* 1997;**26**:155–65.
5. **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.
6. **Gunther KP,** Sun Y. Reliability of radiographic assessment in hip and knee osteoarthritis. *Osteoarthritis Cartilage* 1999;**7**:239–46.
7. **Kellgren JH,** Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;**16**:494–502.
8. **Croft P,** Cooper C, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol* 1990;**132**:514–22.
9. **Resnick D,** Niwayama G. Degenerative disease of extraspinal locations. *Diagnosis of bone and joint disorders*. 2nd ed. Philadelphia: Saunders, 1988:1426–42.
10. **Altman RD.** The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.
11. **Odding E,** Valkenburg HA, Algra D, Vandenouwendland FA, Grobbee DE, 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.
12. **Bierma-Zeinstra SMA,** Bohnen AM, Gincai 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.
13. **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**:1911–18.
14. **Nevitt MC,** Lane NE, Scott JC, Hochberg MC, Pressman AR, Genant HK, *et al*. Radiographic osteoarthritis of the hip and bone mineral density. *Arthritis Rheum* 1995;**38**:907–16.
15. **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.
16. **Felson DT,** Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. *Baillieres Clin Rheumatol* 1995;**9**:253–66.
17. **Bierma-Zeinstra SMA,** Verhagen AP, Berger MY. Het ontwikkelen van classificatiesystemen bij aandoeningen van het bewegingsapparaat. *Huisarts en wetenschap* 2000;**43**:473–7.
18. **Haynes RB,** Wilczynski N, McKibbon A, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in Medline. *J Am Med Informatics Assoc* 1994;**1**:447–58.
19. **Ingvarsson T,** Hagglund G, Lindberg H, Lohmander LS. Assessment of primary hip osteoarthritis: comparison of radiographic methods using colon radiographs. *Ann Rheum Dis* 2000;**59**:650–3.
20. **Birrell F,** Croft P, Cooper C, Hosie G, MacFarlane G, Silman A. Predicting radiographic hip osteoarthritis from range of movement. *Rheumatology (Oxford)* 2001;**40**:506–12.

- 21 **MacGregor AJ**, Antoniadou 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.
- 22 **Antoniades L**, MacGregor AJ, Matson M, Spector TD. A cotwin control study of the relationship between hip osteoarthritis and bone mineral density. *Arthritis Rheumatism* 2000;**43**:1450–5.
- 23 **Bellamy N**, Klestov A, Muirden K, Kuhnert P, Do KA, O’Gorman L, et al. Perceptual variation in categorizing individuals according to American College of Rheumatology classification criteria for hand, knee, and hip osteoarthritis (OA): observations based on an Australian Twin Registry study of OA. *J Rheumatol* 1999;**26**:2654–8.
- 24 **Dickersin K**, Scherer R, Lefebvre C. Systematic reviews: identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.
- 25 **Dickersin K**. Systematic reviews in epidemiology: why are we so far behind? *Int J Epidemiol* 2002;**31**:6–12.
- 26 **Byrt T**, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;**46**:423–9.
- 27 **Smith RW**, Egger P, Coggon D, Cawley MID, Cooper C. Osteoarthritis of the hip joint and acetabular dysplasia in women. *Ann Rheum Dis* 1995;**54**:179–81.
- 28 **Hart DJ**, Spector TD. The classification and assessment of osteoarthritis. *Baillieres Clin Rheumatol* 1995;**9**:407–33.
- 29 **Hart DJ**, Spector TD. Radiographic criteria for epidemiologic studies of osteoarthritis. *J Rheumatol* 1995;**43**(suppl 22):46–8.
- 30 **Spector TD**, Cooper C. Radiographic assessment of osteoarthritis in population studies: whither Kellgren and Lawrence? *Osteoarthritis Cartilage* 1993;**1**:203–6.

## UNUSUAL AND MEMORABLE .....

### Case Number 29: Hitting three with one strike: rapid improvement of psoriatic arthritis, psoriatic erythroderma, and secondary renal amyloidosis by treatment with infliximab (Remicade)

Series editor: Gary D Wright

*Ann Rheum Dis* 2004;**63**:232. doi: 10.1136/ard.2003.014191

A 64 year old woman presented with a 30 year history of psoriasis vulgaris and polyarticular psoriatic arthritis as well as a newly diagnosed renal insufficiency. At the physical examination she showed severe psoriatic erythroderma (fig 1A) and signs of destructive polyarthritis of the joints of the fingers and feet. Radiology disclosed almost complete ankylosis of both wrists, and high grade erosive destruction of the small joints of fingers and toes. Serum creatinine was 330  $\mu\text{mol/l}$  and blood urea nitrogen (BUN) 40 mmol/l. Urine examinations showed a proteinuria of 8 g/24 h (normal value <0.1). Renal biopsy revealed severe AA amyloidosis of the kidneys as the cause of renal insufficiency and nephrotic syndrome. With the exception of ibuprofen she had not so far received regular treatment of her disease.

After failure of low dose oral prednisone, we started treatment with 200 mg infliximab (Remicade) intravenously (3.3 mg/kg body weight). Already after the first infusion she felt strong relief of the joint pain, reduced joint swelling and stiffness, and the erythroderma disappeared (fig 1B). When serum and urine parameters were re-evaluated after the second infusion, the serum creatinine had declined to 280

$\mu\text{mol/l}$  and BUN to 25 mmol/l and the proteinuria was reduced to 0.3 g/24 h. Infusions were repeated after 4 weeks and then every 8 weeks. Interruption of the treatment after 7 months because of a planned operation resulted in a flare of the disease. When the infusions with infliximab were started again, she once more showed a good response. Measures of kidney amyloidosis only slightly worsened when re-evaluated after 9 months (creatinine in the serum 325  $\mu\text{mol/l}$ , BUN 40.5 mmol/l, proteinuria 0.55 g/24 h) as shown by the last follow up.

**C Fiehn**

Department of Internal Medicine V, University Hospital of Heidelberg, Germany

**K Andrassy**

Department of Internal Medicine I, Section of Nephrology, University Hospital of Heidelberg, Germany

Correspondence to: Dr C Fiehn, Department of Internal Medicine V, University of Heidelberg, Hospitalstr. 3, 69123 Heidelberg, Germany; christoph\_fiehn@med.uni-heidelberg.de



**Figure 1** Psoriatic erythroderma (A) before the onset of treatment with 200 mg (3.3 mg/kg) infliximab (Remicade) intravenously and (B) after two infusions.