

# Avascular necrosis of the femoral head: inter- and intraobserver variations of Ficat and ARCO classifications

Marcus Schmitt-Sody · Chlodwig Kirchhoff ·  
Wolfgang Mayer · Michael Goebel · Volkmar Jansson

Received: 6 December 2006 / Revised: 12 December 2006 / Accepted: 16 December 2006 / Published online: 30 March 2007  
© Springer-Verlag 2007

**Abstract** Diagnosis of avascular necrosis (AVN) of the femoral head depends on the combination of clinical symptoms and evaluation of radiographs and/or magnetic resonance imaging (MRI). To evaluate the evolution of AVN, the Ficat and the Association Research Circulation Osseous (ARCO) classification are commonly used to assess both imaging modalities. For comparison reasons, these classifications need to be reliable and reproducible to provide sufficient therapy options for the patient. Therefore, the aim of our study was to evaluate the interobserver reliability and the intraobserver reproducibility of these classifications. Patients with suspected AVN were examined using either radiographs or radiographs and MRI. The radiographs and/or MR images were reviewed initially and at 3 months by two general orthopaedic surgeons, two orthopaedic residents, and two general radiologists using the Ficat classification for radiographs and MR images as well as the ARCO classification for MR images only. In all, 38 patients (54 hips) were

enrolled. There were 10 patients who presented with radiographs and 28 patients with radiographs and MR scans. Paired comparisons revealed a mean interobserver kappa reliability coefficient of 0.39 for the first and of 0.32 for the second review using the Ficat classification for radiographs, whereas for the MR images a mean of 0.39 in the first and of 0.34 in the second reading resulted. The MRI evaluation using the ARCO classification resulted in a mean interobserver reliability coefficient of 0.37 in the first and of 0.31 in the second reading. The mean kappa value for intraobserver reproducibility using the Ficat classification was 0.52 for radiographs and 0.50 for MR images, whereas a reproducibility of 0.43 resulted for the ARCO classification. This study showed poor interobserver reliability and fair intraobserver variability, diminishing any meaningful comparison of studies using the Ficat as well as the ARCO classification. Thus, the Ficat and ARCO staging systems are still not sufficient to reliably assess the status of AVN alone.

Marcus Schmitt-Sody and Chlodwig Kirchhoff contributed equally to this work.

M. Schmitt-Sody (✉) · W. Mayer · V. Jansson  
Department of Orthopedic Surgery,  
University Hospital Grosshadern,  
Marchioninstrasse 15,  
81377 Munich, Germany  
e-mail: mschmitt@med.uni-muenchen.de

C. Kirchhoff  
Department of Orthopedic Surgery and Traumatology,  
University Hospital Innenstadt,  
Nussbaumstrasse 20,  
80336 Munich, Germany

M. Goebel  
Department of Orthopedic Surgery, Bogenhausen Hospital,  
Englschalkingerstrasse 77,  
81925 Munich, Germany

**Résumé** Le diagnostic des nécroses vasculaires de la tête fémorale (AVN) dépend de l'analyse des signes cliniques et de l'évaluation des signes radiographiques ainsi que de l'IRM. Afin d'évaluer l'évolution de l'ostéonécrose de la hanche selon la classification de la Ficat et la classification ARCO, ces examens d'imagerie sont habituellement utilisés. Le but de notre étude est d'évaluer la fiabilité de ces différents paramètres inter et intra observateurs. Matériel et méthode : les patients suspects d'ostéonécrose de la hanche ont été examinés sur le plan radiographique et sur le plan IRM. Les radiographies et les images IRM ont été évaluées initialement trois mois après par deux chirurgiens orthopédistes, deux résidents en orthopédie et deux radiologues en utilisant la classification du Ficat et la classification ARCO pour les images IRM. Résultats : 38 patients (54 hanches) ont été inclus dans cette étude. 10 patients n'ont bénéficié

que d'une radiographie et 28 patients d'une radiographie et d'un scanner. En inter observateur, la classification de Ficat montrait une fiabilité de 0.39 et de 0.32, les images en résonance magnétique montrant une fiabilité inter observateur de 0.39 et de 0.34. L'évaluation par l'IRM en utilisant la classification ARCO montrait une fiabilité inter observateur de 0.37 dans la première lecture et de 0.31 lors de la seconde lecture. La fiabilité intra observateur était de 0.52 en utilisant la classification de Ficat, de 0.50 pour le scanner et de 0.43 pour la classification ARCO. En conclusion : cette étude montre une fiabilité inter observateur faible, une bonne fiabilité intra observateur en utilisant la classification de Ficat et la classification ARCO. Cependant, les classifications ne sont pas suffisantes pour avoir une idée sûre du stade de la nécrose avasculaire.

## Introduction

Osteonecrosis affects both compact and cancellous bone in a circumscribed area. The clinical picture of non-traumatic femoral head necrosis was first described more than 50 years ago. Nowadays avascular necrosis (AVN) continues to present a challenging clinical problem because it affects mostly middle-aged patients in their active phase of living. Another aggravating factor is that in 30–70% of the patients both hips are involved, with a peak incidence at 40 years of age [8]. In contrast to traumatic causes such as subcapital fracture or hip dislocation leading to vascular disruption and acutely deficient perfusion of the femoral head, the exact aetiology of AVN remains obscure [3, 13]. A multifactorial genesis is discussed involving several underlying diseases, such as Gaucher's disease, ionizing radiation, steroid therapy, and risk factors such as alcoholic excess, hyperuricemia, pancreatitis and pregnancy [7, 10]. Osseous necrosis can develop from failure of arterial supply, obstruction of venous drainage, intraluminal capillary obstruction and compression of the capillaries in the bone marrow space.

Current pathophysiological models claim recurrent ischaemic attacks on bone are followed by an increase of intraosseous pressure, most probably due to oedema since bone marrow is functionally a closed compartment. The intraosseous venules and capillaries are consequently compromised, resulting in the vicious circle similar to the compartment syndrome of the extremities.

In the past, several staging classifications have been introduced and used, some based only on plain radiographs and others on imaging modalities such as computed tomography, magnetic resonance imaging (MRI) or bone scintigraphy. It has been demonstrated that an early diagnosis of AVN plays a crucial role with respect to prognosis and therapeutic success [15].

The classification system introduced by Ficat [4] is possibly the most commonly used. However, in the literature this classification system has been criticised because of great inter- as well as intraobserver variability [11, 18]. Moreover, the Ficat classification does not take the size and location of the necrotic area into account. For appropriate consideration of the missing parameters the ARCO (Association Research Circulation Osseous) classification system was introduced [6].

Many authors have recommended treatment on the basis of the symptoms together with the Ficat and the ARCO classification using radiographs and MRI.

Treatment with pulsed electromagnetic fields or core decompression or both, bone grafting and decompression, and rotational transtrochanteric osteotomy have been suggested for Ficat stage I lesions [1, 14, 17]. These procedures as well as vascularised fibular grafting, rotational transtrochanteric osteotomy and intertrochanteric osteotomy have been recommended for stages II A and B [1, 9, 16]. All these treatments as well as surface replacement and total hip replacement have been recommended to treat stage III lesions.

The choice of treatment and the judgement of its efficacy often are directly based on the Ficat/ARCO stage. Thus, the determination of the Ficat or ARCO stage has important consequences as it has a direct bearing on the patient's clinical course.

Although these radiological parameters are commonly used to evaluate the stage and progress of disease, most scores and calculations have not been analysed in more detail. A classification system provides a description of meaningful biological information and should be reproducible from one observer to the next as well as by one observer on separate occasions. The absence of reproducibility clouds the comprehension and comparison of studies and treatment recommendations which are based on such classifications systems.

Therefore, the aim of our study was to evaluate inter- as well as intraobserver reliability and variability of commonly used parameters of Ficat and ARCO on plain radiographs and MR images for the classification of femoral head necrosis.

## Patients and methods

Between 1998 and 2004, all patients with suspected AVN were examined at our University Hospital's department of radiology using only plain radiographs or both radiographs and MRI. The inclusion criteria for suspected AVN were typical clinical symptoms such as sudden pain in the hip following trauma or, in the absence of trauma, the coincidence of certain co-morbidities such as Gaucher's disease, diabetes, hypertension, rheumatic diseases, etc.

All identifying data on either radiographs or MR images were anonymised and randomised. Each radiograph and MR sequence was reviewed and classified by six observers: two general orthopaedic surgeons, two orthopaedic residents and two general radiologists. All observers were familiar with the Ficat as well as the ARCO classification system and had used it in daily clinical routine previously. The observers were provided with a copy of both classification systems to be used. They were allowed to refer to the copies as often as necessary during reading of the images. Each observer read the images on his own apart from a third person not involved in the study, documenting the results. During or after the reading, no discussion among the observers was allowed. The observers were given as much time as they needed to carefully review each radiograph or MR-image sequence. After a decision had been made, the images of the next hip/hips were displayed until all hips had been classified and documented.

After a period of three months all hips were reviewed again on a second occasion. In the interim the images were not available for any observer, no feedback regarding the initial reading was provided. The second review of images was performed in the same way as the first image reading session.

#### Statistical analysis

Computer-assisted statistical analysis (SAS, Heidelberg, Germany) was used to determine the inter- and intra-observer reliability of both classification systems for radiographs and MRI by calculating the weighted Cohen's kappa index. The kappa values were generated by setting the observed proportion of agreement in relation to the proportion of agreement by chance. Kappa values of less than 0.5 indicated poor agreement and values greater than 0.75 were considered as excellent agreement. The accuracy or the measurement of how close an experimental observation lies to the true value was impossible because the correct classification for each evaluated hip was not available and not known. Therefore, the level of agreement between observers in terms of interobserver reliability and between each review of each observer in terms of intra-observer variability was assessed over time for both the Ficat and the ARCO classification system.

#### Results

Overall 38 patients (16 women, 22 men, mean age 55.5 years, SD 10.6 years) were enrolled; 54 hips were included in this study and evaluated.

Ficat classification (see Table 1)

##### *Interobserver reliability*

For the interobserver reliability of radiographs (Table 1), a mean reliability coefficient of 0.36 resulted (range 0.11 to 0.68) averaged over both reviews. For the first review, a mean interobserver reliability coefficient of 0.39 resulted, whereas for the second review the mean interobserver reliability coefficient was 0.32.

The MR evaluation revealed a mean interobserver reliability coefficient of 0.37 (range 0.23 to 0.70). The mean interobserver reliability coefficient was 0.39 for the first review and 0.34 for the second review.

##### *Intraobserver reproducibility*

The mean weighted kappa intraobserver reproducibility coefficient for radiographs was 0.53 (range 0.29 to 0.76) among the six observers.

For the evaluation of the MR images a mean weighted kappa intraobserver reproducibility coefficient of 0.50 resulted (range 0.29 to 0.71).

ARCO classification (see Table 2)

##### *Interobserver reliability*

For the interobserver reliability of MRI (Table 2), a mean reliability coefficient of 0.35 resulted (range 0.06 to 0.56), averaged over both reviews. For the first review, a mean interobserver reliability coefficient of 0.38 resulted, whereas for the second review the mean interobserver reliability coefficient was 0.31.

##### *Intraobserver reproducibility*

For the evaluation of the MR images a mean weighted kappa intraobserver reproducibility coefficient of 0.44 resulted (range 0.26 to 0.56).

#### Discussion

In this study the intra- as well as the interobserver variability and reproducibility for the evaluation of avascular necrosis of the femoral head was assessed using the Ficat as well as the ARCO classification. We demonstrated a poor interobserver reliability and a fair intraobserver variability, diminishing any meaningful comparison of studies using the Ficat as well as the ARCO classification system. Thus, these staging systems are still not sufficiently

**Table 1** Scheme of Ficat classification (1985) [4]

Stage	Radiographic signs	Clinical features
0	Inconspicuous/normal findings	0 (“silent hip”)
I	Inconspicuous findings or minor changes (slight patchy osteoporosis, blurring of trabecular pattern, subtle loss of clarity)	+
II A	Diffuse/ focal radiological changes (osteoporosis, sclerosis, cysts)	+
II B	Subchondral fracture (“crescent sign”) segmental flattening of femoral head (“out-of-round appearance”)	+
III	Broken contour of femoral head, bone sequestrum, joint space normal	++
IV	Flattened contour of femoral head, decreased joint space collapse of femoral head, acetabular osteoarthrotic changes	+++

reliable to assess the status of avascular necrosis of the hip on a use-alone basis.

Radiological evaluation of the stage and extent of disease plays an important role in avascular necrosis of the femoral head [15]. In the literature an increasing number of studies deal with the predictive quality rating of radiological classification schemes [15, 20].

However, the reproducibility of these analysis techniques has not been evaluated. These orthopaedic classification systems provide subdivisions in the spectrum of presentation of certain disease processes. These classification systems need to be of reproducible character among different observers as well as by the same observer on different occasions.

Sophisticated diagnostics should provide the information as to whether joint-conserving surgery is reasonable or already too late. Thus, diagnostic instruments providing high reliability are necessary especially for the correct and adequate staging of femoral head necrosis.

The modified Ficat system described by Smith et al. [18] does not provide sufficient information for the staging of the disease. In this study, two different classification systems, the Ficat as well as the ARCO classification, were used to stage osteonecrosis of the hip using both radiographs and MR images. Other orthopaedic studies dealing with evaluation of orthopaedic classifications of osteonec-

rosis of other bones than the hip also used kappa statistics, describing disappointing results with kappa values ranging from 0.4 to 0.57 for interobserver variability and from 0.58 to 0.69 for intraobserver reproducibility [2, 5]. A modified scheme of kappa values, as recommended by Smith et al. [18] was used for the interpretation of this study’s results, interpreting kappa values between 0.50 and 0.75 as fair and >0.75 as excellent. According to these modified guidelines, the present study presented poor interobserver reliability using the Ficat classification for radiographs and MRI in addition to the ARCO classification for MRI. Fair intraobserver variability was found for the Ficat classification on radiographs and MRI compared to poor variability within the readings of each observer using the ARCO classification. No excellent results were found regarding interobserver reliability or intraobserver variability. Our findings reinforce the results of Smith et al. who used the modified Ficat classification only.

Although not statistically analysed in this study, the reviewers were least likely to change their classification when initially classifying hips as stage I or IV. This is presumably because stage I or IV present more clearly obvious changes either of normal or severely osteonecrotic character respectively. In comparison, in most classifications of stage II A/B or III these hips were differently staged in the second review, most likely due to various

**Table 2** Scheme of ARCO classification system (1992) [6]

Stage	Radiological findings	Subclassification
0	positive: histology negative/normal: Radiograph/CT/MRI/scintigraphy	–
I	positive: MRI and/or bone scintigraphy negative/normal: radiograph/ CT	+’ <sup>(a)</sup>
II	Radiograph: sclerotic, cystic or osteoporotic changes of femoral head	+’ <sup>(a)</sup>
III	Radiograph: subchondral fracture (“crescent sign”)	+’ <sup>(a)</sup>
IV	Radiograph: flattening of femoral head	++’ <sup>(b)</sup>
V	Radiograph: flattening of femoral head and osteoarthrotic changes: decreased joint space and acetabular changes	++’ <sup>(b)</sup>
VI	Complete joint destruction	–

<sup>(a)</sup> Location of femoral head necrosis: 1) medial third, 2) median third, 3) lateral third. Size of femoral head necrosis: A) <15%, B) 15–30%, C) >30%

<sup>(b)</sup> Intrusion degree of femoral head contour: A) <2 mm, B) 2–4 mm, C) >4 mm

interpretations of the phrases of both classifications with regard to the description of middle stages [4].

As stated above, the reliability of our results showed a high variability of the Ficat classification. Thus, it was concluded that the Ficat classification is not appropriate for the evaluation of femoral head necrosis and more reliable classifications should be used. The ARCO scheme was actually developed to be able to simultaneously evaluate different aspects of femoral head necrosis. Unfortunately, in comparison to the Ficat classification no significant improvement of reliability resulted from using the ARCO classification.

Due to the lack of excellent interobserver reliability results it is not possible to make comparisons between studies of different study centres even in cases using similar classification. Also, because no excellent intraobserver variability was described it is not plausible to rely on outcome studies when they are based on either radiographs or MRI since both classifications did not result in excellent intraobserver variability, being assessed over time, e.g. before and after treatment by the same observer or group of observers.

So far only one study exists describing the possibility of radiologically assessing the extent of necrotic area of the femoral head [19]. Unfortunately, this study presents only a small number of patients and is limited to the initial necrotic stages; thus, the figures shown are not sufficiently representative. Although an exact and precise evaluation of the extent of affected area should be possible using MRI and CT, it remains unclear whether these modalities provide exact information on the real necrotic volume [12, 20], and whether a higher accuracy of diagnosis would inevitably lead to improved treatment results.

Unfortunately, the classification schemes for osteonecrosis of the femoral head of Ficat and ARCO do not have acceptable interobserver reliability and intraobserver reproducibility on which treatment protocols and determination of outcome can be based. However, MRI and CT offer a more detailed view of the involvement of the femoral head, subchondral collapse, narrowing of the joint space and acetabular changes found with the progression of this disease. Nevertheless radiographs will continue to play an important role in the evaluation of follow-up of the disease, even though other, modern imaging modalities such as MRI are available.

**Acknowledgments** We cordially have to thank Ms. Andrea van Dyke for her substantial work regarding the evaluation of the radiographs and MRI data. The presented data form part of her doctoral thesis. Moreover we have to thank the Institute of Medical Biometry and Epidemiology of the Ludwig-Maximilian-University Munich for their support regarding the statistical analysis.

## References

1. Aaron RK, Lennox D, Bunce GE, Ebert T (1989) The conservative treatment of osteonecrosis of the femoral head: a comparison of core decompression and pulsing electromagnetic fields. *Clin Orthop* 249:209–218
2. Andersen E, Jorgensen LG, Heddam LT (1990) Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. *Injury* 21:377–378
3. Catto M (1965) A histological study of avascular necrosis of the femoral head after transcervical fractures. *J Bone Joint Surg Br* 47:749–776
4. Ficat RP (1985) Idiopathic bone necrosis of the femoral head. Early diagnosis and treatment. *J Bone Joint Surg Br* 67:3–9
5. Frandsen PA, Anderson E, Madsen F, Skjodt T (1988) Garden's classification of femoral neck fractures. An assessment of interobserver variations. *J Bone Joint Surg* 70:1048–1051
6. Gardeniers JW (1992) A new international classification of osteonecrosis of the ARCO Committee on terminology and classification. *J Jpn Orthop Assoc* 66:18–20
7. Hofmann S, Engel A, Neuhold A, Leder K, Kramer J, Plenk H (1993) Bone-marrow edema syndrome and transient osteoporosis. *J Bone Joint Surg Br* 75:210–216
8. Jacobs B (1978) Epidemiology of traumatic and nontraumatic osteonecrosis. *Clin Orthop* 277:41–45
9. Jacobs MA, Hungerford DS, Krachow KA (1989) Intertrochanteric osteotomy for avascular necrosis of the femoral head. *J Bone Joint Surg* 71-B:200–204
10. Jones JP Jr (1994) Concepts of aetiology and early pathogenesis of osteonecrosis. *Instr Course Lect* 43:499–512
11. Kay RM, Lieberman JR, Dorey FJ, Seeger LL (1994) Inter- and intraobserver variation in staging patients with proven avascular necrosis of the hip. *Clin Orthop* 307:124–129
12. Koo KH, Kim R (1995) Quantifying the extent of osteonecrosis of the femoral head. A new method using MRI. *J Bone Joint Surg Br* 77:875–880
13. Langer R, Scholz A, Langer M, Astinet F, Schetlick G, Ferstl F, Felix R (1991) Superselektive intraarterielle DAS bei Hüftkopfnekrose und Schenkelhalsfraktur. *RöFo* 154:587–592
14. Lennox DW, Murrah RL, Ebert T, Carbone J (1993) The efficacy and safety of core decompression of the hip as a treatment for osteonecrosis. *Complicat Orthop* 47:39–42
15. Mont MA, Jones LC, Pacheco I, Hungerford DS (1998) Radiographic predictors of outcome of core decompression for hips with osteonecrosis stage III. *Clin Orthop* 354:159–168
16. Saito S, Ohzono K, Ono K (1988) Joint-preserving operations for the idiopathic avascular necrosis of the femoral head. *J Bone Joint Surg* 70-B:78–84
17. Smith SW, Fehring TK, Grffin WL, Beaver WB (1995) Core decompression of the osteonecrotic femoral head. *J Bone Joint Surg* 77-A:674–680
18. Smith SW, Meyer RA, Connor PM, Smith SE, Hanley EN Jr (1996) Interobserver reliability and intraobserver variability of the modified Ficat classification system of osteonecrosis of the femoral head. *J Bone Joint Surg Br* 78:1702–1706
19. Steinberg ME, Hayken GD, Steinberg DR (1995) A quantitative system for staging avascular necrosis. *J Bone Joint Surg Br* 77:34–41
20. Steinberg ME, Bands RE, Parry S, Hoffmann E, Chan T, Hartman KM (1999) Does lesion size affect the outcome in avascular necrosis? *Clin Orthop* 367:262–271