The Eftekhar and Kerboul classifications in assessment of developmental dysplasia of the hip in adult patients. Measurement of inter- and intraobserver reliability

Alexander Brunner, MD · Benjamin Ulmar, MD · Heiko Reichel, MD, Ph.D. · Ralf Decking, MD

Received: 11 April 2007/Accepted: 12 November 2007/Published online: 18 December 2007
© Hospital for Special Surgery 2007

Abstract *Aim:* To measure the inter- and intraobserver reliability of two radiologic classification systems in the evaluation of severity of osteoarthritis secondary to developmental dysplasia of the hip in adult patients and to compare these systems to historically published data on two more commonly used classification systems described by Crowe and by Hartofikakidis. *Material and methods:* Eighty-six dysplastic hips on 66 anteroposterior standard pelvic x-rays were rated according to the criteria of Eftekhar and Kerboul by three observers with different levels of clinical training. To assess intraobserver reliability, the measurement was repeated 3 months later. Statistical analysis was performed by calculating the weighted kappa correlation coefficient and the overall kappa coefficient. *Results:* Both classification systems showed a sufficient interobserver reliability, reaching a kappa coefficient of 0.73 for the system according to Eftekhar and 0.716 for the Kerboul classification system. Intraobserver reliability revealed a kappa coefficient of 0.778 for the Eftekhar system and 0.697 for the classification according to Kerboul. *Conclusion:* Both systems showed good inter- and intraobserver reliability for use in daily practice. However, they did not reach the known reliability of the grading systems published by Crowe and by Hartofikakidis. Because these

systems are more frequently used, we would recommend one of them to grade the severity of osteoarthritis secondary to developmental dysplasia of the hip in adult patients instead of the ones described by Eftekhar or by Kerboul.

Key words hip · dysplasia · prosthesis · classification

Introduction

Untreated developmental dysplasia of the hip (DDH) is one of the most common reasons for secondary coxarthrosis in female adult patients [1]. Radiological signs of dysplasia have been described in 25% of all patients being treated by total hip replacement (THR) [2]. The severity of dysplasia may range from slight hypoplasia of the acetabulum with only little lateral displacement of the joint center to an insufficient acetabular roof to contain the femoral head leading to its complete dislocation. In its milder forms with only mild subluxation, the shallow acetabulum commonly has a wide opening and is oval in shape. In cases of a high dislocation of the hip, the femoral head does not articulate with the original acetabulum, which is often rudimentary [3].

Focusing on the femoral side of dysplasia, a coxa valga deformity with posterior translation of the greater trochanter and excessive anteversion of the femoral neck is common in almost all subluxed hip joints. In severe forms of femoral dysplasia, the femur is small with a narrow medullary canal and thin, ecstatic cortical bone [4].

Because of these morphologic abnormalities, THR in the setting of developmental dysplasia of the hip can be technically demanding [4, 5]. When the dysplasia is mild, only minimal modification to the surgical technique of total hip arthroplasty is required [4]. In more severe cases, however, the operation is substantially more complex, and special techniques to restore an anatomic location of the femoral head and to avoid excessive lengthening of the leg and inadvertent sciatic nerve injury are required. Furthermore, with the anatomic femoral abnormalities, special implants may be required to deal with

A. Brunner, MD · B. Ulmar, MD · H. Reichel, MD, Ph.D. · R. Decking, MD (✉)
Department of Orthopaedics,
University of Ulm,
Oberer Eselsberg 45, Ulm, 89081, Germany
e-mail: ralf.decking@uni-ulm.de

A. Brunner, MD
e-mail: a-r.Brunner@web.de

B. Ulmar, MD
e-mail: benjamin.ulmar@uni-ulm.de

H. Reichel, Ph.D.
e-mail: heiko.reichel@uniklinik-ulm.de

them [4, 6]. A number of studies showed [7–13] that THR in the setting of developmental dysplasia has higher rates of complications and inferior outcomes when compared with THR performed for primary osteoarthritis. Moreover, the results depend on the severity of dysplasia, and patients with lower grades of dysplasia tend to have better results than patients with higher grades of dysplasia [14], so interpretation of long-term results after THR requires detailed information about the degree of dysplasia.

A number of classification systems [15–19] have been proposed to describe and grade the anatomic abnormalities

and the severity of dysplastic dislocated hips. In 1978, Eftekhar [16] defined a radiologic classification system depending on the grade of dislocation of the femoral head. The severity of hip dysplasia was divided into four stages, ranging from dysplasia to complete dislocation (Fig. 1).

In type A, the acetabulum is slightly elongated and appears dysplastic, accommodating a flattened mushroom-shaped femoral head. Eftekhar described type B and C as intermediate and high dislocations, respectively. The true acetabulum is poorly developed, but the floor is thick and easily identifiable following removal of fibrofatty tissue from original site of true acetabulum. The lower border of false acetabulum identifies roof of original acetabulum. Type D: old, unreduced dislocation. Head has never been in contact with ilium. There is no pseudoacetabulum; the original site of acetabulum can hardly be recognized

Fig. 1 The Eftekhar classification. *Type A*: slightly elongated dysplastic acetabulum accommodating a flattened mushroom-shaped femoral head. *Types B and C*: intermediate and high dislocation. The true acetabulum is poorly developed, but floor is thick and easily identifiable following removal of fibrofatty tissue from original site of true acetabulum. The lower border of false acetabulum identifies roof of original acetabulum. *Type D*: old, unreduced dislocation. Head has never been in contact with ilium. There is no pseudoacetabulum; the original site of acetabulum can hardly be recognized

from the site of the true acetabulum. The lower border of the false acetabulum identifies the roof of the original acetabulum.

Type D describes an old, unreduced dislocation. The head actually has never been in contact with the ilium, so there is no pseudoacetabulum. The original site of the acetabulum can hardly be recognized. It only represents a narrow “isthmus” at the site of the old triradiate cartilage.

In 1987 Kerboul, Mathieu, and Sauzies [18] used the anteroposterior position of the femoral head to distinguish between three different forms of dysplasia and dislocation (Fig. 2). An anterior dislocation where the femoral head is

located in front of the original acetabulum is defined as type A. Type B describes an intermediate dislocation. The femoral head articulates with the ilium at the same anteroposterior level as the original acetabulum. In type C, the femoral head is dislocated behind the original acetabulum (posterior dislocation).

A number of studies [16, 18, 20–22] have used the classifications of Eftehkar and Kerboul to describe the preoperative anatomic situation before THR. The use of any radiologic classification system requires sufficient reliability, as measured by inter- and intraobserver correlation. Unfortunately, the measurement of reliability was not part

Fig. 2 The Kerboul classification. *Type A*: anterior dislocation where the femoral head is located in front of the original acetabulum. *Type B*: intermediate dislocation. The femoral head articulates with the ilium at the same anteroposterior level as the original acetabulum. *Type C*: posterior dislocation. The femoral head is dislocated behind the original acetabulum

of the original publications of Eftekhar and Kerboul [16, 18], and we have not been able to find any studies evaluating it. This study was performed to measure the inter- and intraobserver reliability of these two radiologic classification systems, to evaluate the comparability of data presented in the literature, and to see if these systems offer an advantage over other classifications used.

Material and methods

From the databases of two orthopedic hospitals (Orthopaedic Department of the University of Ulm, Germany, and the Orthopaedic Department of St. Francis Hospital, Muenster, Germany), a total of 96 patients within three consecutive years were identified with the preoperative diagnosis “coxarthrosis secondary to developmental dysplasia of the hip” who had been scheduled for THR. In 69 of these patients (48 women and 21 men), preoperative anteroposterior pelvis x-ray films could be obtained. Twenty of the these 69 patients showed a bilateral dislocation. We reviewed the preoperative anteroposterior pelvic x-ray films of these patients. In three patients, the x-ray films were underexposed and unsuitable for evaluation, which left 86 dislocated hips on 66 anteroposterior pelvic x-ray films (46 women and 20 men) for review and analysis. The mean age of these patients was 54 years (range: 17–83 years).

To consider the possible effect of the observers' experience on classifying the hips, we chose three observers with different levels of orthopedic training and experience (one fellowship-trained consultant specializing in joint

replacement, one fourth-year resident, and one first-year resident). The three observers made themselves familiar with the original publications of Eftekhar and Kerboul [16, 18]. Prior to the study, the observers were trained on the nuances and subtleties of the classification systems by reviewing ten x-rays as a group. In accordance with HIPPA regulations, all identifying marks were removed from the films, which were then sorted, labeled, and given to the observers in a random order by one of the investigators who was not an observer. All x-rays were graded by each observer in separate rooms and in random order first according to the Eftekhar system and later according to the criteria of Kerboul et al. Each observer reviewed the x-rays again 3 months later. The order of the x-rays was randomized again to prevent possible recall.

Statistical analysis was performed by calculating the weighted kappa correlation coefficient according to Cicchelli and Allison [23], as well as the overall kappa coefficient according to Fleiss [24]. Interpretation of the kappa coefficients was performed using the criteria of Landis and Koch [25]. They defined a kappa coefficient of more than 0.8 as excellent, between 0.6 and 0.8 as good (exceeding chance), between 0.4 and 0.6 as moderate, and less than 0.4 as poor correlation.

Results

Measurement of interobserver reliability between two observers (Table 1) showed kappa coefficients between 0.608 and 0.803 (mean: 0.73) for the Eftekhar system and

Table 1. Interobserver reliability of the Eftekhar and Kerboul classification between two observers [weighted kappa coefficient (κ) with 95% confidence intervals and p values for the differences between the calculated kappas]

n = 86	(κ) A/B	(κ) A/C	(κ) B/C
Eftekhar t1	0.608 ($\pm$ 0.11)	0.726 ($\pm$ 0.1)	0.722 ($\pm$ 0.097)
p-value	A/B-A/C = 0,078		A/C-B/C = 0,27
	A/B-B/C = 0,072		
Eftekhar t2	0.723 ($\pm$ 0.1)	0.803 ($\pm$ 0.092)	0.802 ($\pm$ 0.083)
p-value	A/B-A/C = 0,084		A/C-B/C = 0,21
	A/B-B/C = 0,3		
Kerboul t1	0.704 ($\pm$ 0.12)	0.698 ($\pm$ 0.126)	0.819 ($\pm$ 0.1)
p-value	A/B-A/C = 0,25		A/C-B/C = 0,07
	A/B-B/C =		
Kerboul t2	0.682 ($\pm$ 0.138)	0.605 ($\pm$ 0.148)	0.79 ($\pm$ 0.115)
p-value	A/B-A/C = 0,12		A/C-B/C = 0,08
	A/B-B/C = 0,072		

Intrater reliability was calculated after two separate readings (t1, t2) of the three observers, A = orthopedic consultant, B = experienced orthopedic resident, C = first-year resident

Table 2. Interobserver reliability of the Eftekhar and Kerboul classification between all three observers together [weighted overall kappa coefficient (κ) with 95% confidence intervals and p values for the differences between the calculated kappas]

n = 86	(κ) A/B/C	p-value		
Eftekhar t1	0.536 ($\pm$ 0.076)	0,07		0,09
Eftekhar t2	0.668 ($\pm$ 0.076)		0,27	
Kerboul t1	0.673 ($\pm$ 0.093)	0,16		0,09
Kerboul t2	0.658 ($\pm$ 0.094)			

Overall interrater reliability was calculated after two separate readings (t1; t2) of the three observers, A = orthopedic consultant, B = experienced orthopedic resident, C = first-year resident

between 0.605 and 0.819 (mean: 0.716) for the Kerboul classification. According to the criteria of Landis and Koch, the Eftekhar system showed an “excellent” interobserver correlation in two and a “good” correlation in four cases. The Kerboul classification showed an “excellent” interobserver correlation in one and “good” correlation in five cases. Using the interpretation of Landis and Koch, the interobserver reliability between all three observers (Table 2) had one “good” and one “moderate” correlation for the Eftekhar system and two “good” correlations for the Kerboul classification.

Intraobserver reliability (Table 3) revealed kappa coefficients between 0.717 and 0.853 (mean, 0.778) for the Eftekhar system and between 0.643 and 0.733 (mean, 0.697) for the Kerboul classification. Intraobserver reliability was “excellent” in one and “good” in two cases for the Eftekhar system and “good” in all three cases for the Kerboul classification. The Eftekhar system showed a slight increase of kappa coefficients in the second reading, which was not observed in the Kerboul classification. All kappa coefficients showed statistical significance revealing p values between 0.001 and 0.01. The differences between the kappa coefficients calculated for the intraobserver and

interobserver reliability ranged from 0.001 to 0.185. The p values for these differences ranged from 0.056 to 0.32 (Tables 1, 2, and 3). As such, with the level of significance set at $\alpha=0.05$, none of the differences between the readings and the observers were significant.

In four cases (4.7%), all three raters found it extremely difficult to distinguish between Eftekhar types B and C in all measurements. In three cases (3.5%), observers found it difficult to differentiate between Kerboul types A and B.

Discussion

These two radiologic classification systems showed sufficient inter- and intraobserver reliability for the use in scientific studies and in daily practice. Overall the Eftekhar classification demonstrated slightly better results than the system according to Kerboul et al.

In addition to reliability, for a classification system to be useful, it needs to be valid. This means that the classification system should provide an accurate measurement of what is actually being measured. In this case, the intraoperative findings would be the gold standard. This study

Table 3. Intraobserver reliability of the Eftekhar and Kerboul classification for each of the three observers [weighted kappa coefficient (κ) with 95% confidence intervals and p values for the differences between the calculated kappas]

n = 86	(κ) A	(κ) B	(κ) C
Eftekhar t1/t2	0.765 ($\pm$ 0.102)	0.717 ($\pm$ 0.095)	0.853 ($\pm$ 0.074)
p-value	A-B = 0,28		B-C = 0,065
	A-C = 0,182		
Kerboul t1/t2	0.733 ($\pm$ 0.126)	0.716 ($\pm$ 0.118)	0.643 ($\pm$ 0.136)
p-value	A-B = 0,318		B-C = 0,292
	A-C = 0,215		

Intrater reliability was calculated between two separate readings (t1, t2) of each of the three observers, A = orthopedic consultant, B = experienced orthopedic resident, C = first-year resident

did not measure the concordance between intraoperative findings and the radiographic ratings. This is a major weakness of this study, but it is also a weakness of the other radiographic classifications used today. To this point, the Eftekhari system has some potential downsides that might limit its validity. It defines an intermediate dislocation of the femoral head as type B and a high dislocation as type C. Some dysplastic hips rated in this study fit neither type B nor type C as proposed by Eftekhari.

Similarly, the Kerboul system uses the anteroposterior position of the femoral head to grade the severity of hip dysplasia. In some cases, the observers found it extremely difficult to locate the anteroposterior displacement of the femoral head by using a standard pelvic x-ray. It might be possible that the “training session” had biased the three observers for an intraobserver-reliability with a mean kappa coefficient of 0.716. Hence, for use in daily practice, a cross-table lateral hip x-ray is necessary. In many hospitals, this is not available for retrospective studies and would furthermore increase the number of x-rays needed when using this classification prospectively. In addition, the anteroposterior location of the femoral head does not tell the surgeon much about the severity of the dysplasia, and may not have much relevance to surgical practice. As reported by several authors [3, 5, 14–17, 26, 27], proximal dislocation of the femur and severity of acetabular dysplasia are important factors that influence preoperative planning, operative procedure, and functional postoperative outcome after THR for DDH. The Kerboul classification gives insufficient information about any of these factors, and we were unable to localize any publication evaluating the influence of the femur’s anteroposterior position on the severity of acetabular dysplasia or the height of dislocation.

Some authors [28, 29] have suggested that clinical experience might affect the interobserver agreement between two observers with different training levels. In the present study, different levels of clinical experience did not influence the interobserver reliability.

A number of other classification systems [15, 17, 19] have been published to grade osteoarthritis secondary to DDH. Most of the international literature is published using either the system of Crowe et al. [15] or the one of Hartofilakidis et al. [17]. Crowe et al. [15] defined a four-stage system classifying the degree of dislocation in terms of the percentage of proximal displacement of the femoral head in relation to the height of the pelvis, resulting in a calculated coefficient, which translates into one of the four types.

Hartofilakidis et al. [17] used the pathology of the dysplastic acetabulum to distinguish between three different types of dysplasia, discriminating between a primary and a secondary acetabulum and the relation of the head in relation to these structures. Both systems have shown a better reproducibility than the systems of Eftekhari and Kerboul et al. when studied with the same protocol [30] used in this study. Interrater kappa-coefficients ranged from 0.75 to 0.88 (Crowe) and 0.68 to 0.84 (Hartofilakidis), and intrarater

coefficients between 0.76 and 0.94 for the Crowe and 0.72 and 0.84 for the Hartofilakidis classification.

Conclusion

The methods of Eftekhari and Kerboul showed a sufficient inter- and intraobserver reliability, yet questions remain regarding their validity and regarding the relevance of the information obtained from the Kerboul classification. Furthermore, both of the systems have been used to a far lesser extent than the classifications by Crowe and Hartofilakidis, which also had a slightly better reproducibility. In addition, there are more data on their correlation regarding the outcome of THA according to the severity of dysplasia. With the classification of Crowe being increasingly popular, especially in the English literature, it may be the best choice for a classification system with the data available today.

References

1. Solomon L, Schnitzler CM (1983) Pathogenetic types of coxarthrosis and implications for treatment. *Arch Orthop Trauma Surg* 101:259–261
2. Gunther KP, Stürmer T, Trepte CT, et al. (1999) Incidence of joint-specific risk factors in patients with advanced cox- and gonarthroses in the Ulm Osteoarthritis Study. *Z Orthop Ihre Grenzgeb* 137:468–473
3. Paavilainen T (1997) Total hip replacement for developmental dysplasia of the hip. *Acta Orthop Scand* 68:77–84
4. Harris WH (1998) Total hip arthroplasty in the management of congenital hip dislocation. In: Callaghan JJ, Rosenberg AG, Rubash HE (eds) *The adult hip*. Lippincott-Raven, Philadelphia, pp 1165–1182
5. Charnley J, Feagin JA (1973) Low-friction arthroplasty in congenital subluxation of the hip. *Clin Orthop* 91:98–113
6. Decking R, Brunner A, Gunther KP, et al. (2006) Total hip arthroplasty in congenital dysplasia of the hip: follow-up of a small-dimensioned, cemented straight stem. *Z Orthop Ihre Grenzgeb* 144:380–385
7. Chandler HP, Reineck FT, Wixson RL, et al. (1981) Total hip replacement in patients younger than thirty years old. A five-year follow-up study. *J Bone Joint Surg Am* 63:1426–1434
8. Collis DK (1984) Cemented total hip replacement in patients who are less than fifty years old. *J Bone Joint Surg Am* 66:353–359
9. Collis DK (1991) Long-term (twelve to eighteen-year) follow-up of cemented total hip replacements in patients who were less than fifty years old. A follow-up note. *J Bone Joint Surg Am* 73:593–597
10. Cornell CN, Ranawat CS (1986) Survivorship analysis of total hip replacements. Results in a series of active patients who were less than fifty-five years old. *J Bone Joint Surg Am* 68:1430–1434
11. Dorr LD, Takei GK, Conaty JP (1983) Total hip arthroplasties in patients less than forty-five years old. *J Bone Joint Surg Am* 65:474–479
12. Halley DK, Wroblewski BM (1986) Long-term results of low-friction arthroplasty in patients 30 years of age or younger. *Clin Orthop* 211:43–50
13. Ivory JP, Kershaw CJ, Choudhry R, et al. (1994) Autophor cementless total hip arthroplasty for osteoarthritis secondary to congenital hip dysplasia. *J Arthroplasty* 9:427–433
14. Chougle A, Hemmady MV, Hodgkinson JP (2005) Severity of hip dysplasia and loosening of the socket in cemented total hip replacement. A long-term follow-up. *J Bone Joint Surg Br* 87:16–20

15. Crowe JF, Mani VJ, Ranawat CS (1979) Total hip replacement in congenital dislocation and dysplasia of the hip. *J Bone Joint Surg Am* 61:15–23
16. Eftekhari N (1978) Principles of total hip arthroplasty. C V Mosby, St. Louis, pp 437–455
17. Hartofilakidis G, Stamos K, Ioannidis TT (1988) Low friction arthroplasty for old untreated congenital dislocation of the hip. *J Bone Joint Surg Br* 70:182–186
18. Kerboul M, Mathieu M, Sauzières P (1987) Total hip replacement for congenital dislocation of the hip. In: Postel M, Kerboul M, Evrard J, Courpié JP (eds) Total hip replacement. Springer, Berlin Heidelberg New York, pp 51–66
19. Mendes DG, Said MS, Aslan K (1996) Classification of adult congenital hip dysplasia for total hip arthroplasty. *Orthopedics* 19:881–887
20. Anderson MJ, Harris WH (1999) Total hip arthroplasty with insertion of the acetabular component without cement in hips with total congenital dislocation or marked congenital dysplasia. *J Bone Joint Surg Am* 81:347–354
21. Kerboul M. (1989) Implantation of a total prosthesis in the deformed hip—exemplified by congenital hip dislocation. *Orthopade* 18:397–417
22. Knecht A, Witzleb WC, Beichler T, et al. (2004) Functional results after surface replacement of the hip: comparison between dysplasia and idiopathic osteoarthritis. *Z Orthop Ihre Grenzgeb* 142:279–285
23. Cicchelli D, Allison T (1971) A new procedure for assessing reliability of scoring EEG sleep recordings. *Am J EEG Technol* 11:101–109
24. Fleiss J (1981) Statistical methods for rates and proportions. Wiley, New York
25. Landis J, Koch G (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
26. Becker DA, Gustilo RB (1995) Double-chevron subtrochanteric shortening derotational femoral osteotomy combined with total hip arthroplasty for the treatment of complete congenital dislocation of the hip in the adult. Preliminary report and description of a new surgical technique. *J Arthroplasty* 10:313–318
27. Cameron HU, Botsford DJ, Park YS (1996) Influence of the Crowe rating on the outcome of total hip arthroplasty in congenital hip dysplasia. *J Arthroplasty* 11:582–587
28. Rasmussen S, Madsen PV, Bennicke K (1993) Observer variation in the Lauge–Hansen classification of ankle fractures. Precision improved by instruction. *Acta Orthop Scand* 64:693–694
29. Sidor ML, Zuckerman JD, Lyon T, et al. (1993) The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg Am* 75:1745–1750
30. Decking R, Brunner A, Decking J, et al. (2006) Reliability of the Crowe und Hartofilakidis classifications used in the assessment of the adult dysplastic hip. *Skeletal Radiol* 35:282–287