

Reliability and validity of the Hartofilakidis classification system of congenital hip disease in adults

C. K. Yiannakopoulos · T. Xenakis · T. Karachalios ·
G. C. Babis · G. Hartofilakidis

Received: 29 August 2007 / Revised: 9 September 2007 / Accepted: 18 September 2007
© Springer-Verlag 2007

Abstract The reliability and validity of the Hartofilakidis et al. classification system in adults with congenital hip disease (CHD) were examined. The radiographs of 102 adult patients (158 hips) with CHD were independently assessed by three senior surgeons. Interobserver variability was assessed by examining the agreement between the three raters while validity of the classification system was assessed by examining the agreement between the assessment by either one of the three raters and the intraoperative finding (reference standard). The interobserver agreement between the three observers was high ranging from 0.720 to 0.854 (substantial to excellent) while the agreement of the preoperative prediction with the intraoperative findings was 87.4% ($K=0.823$, excellent agreement). The Hartofilakidis et al. classification system reliably predicts from preoperative pelvis radiographs the bone deficiencies encountered during the operation.

Résumé Nous avons voulu tester dans ce travail la fiabilité et la viabilité de la classification d'Hartofilakidis dans les séquelles des affections congénitales de la hanche. Les radios de 102 adultes (158 hanches) avec lésions congénitales ont été analysées, de façon indépendante par trois chirurgiens seniors avec validation inter observateur et intra observateur. Les résultats de cette analyse inter observateur sont hautement significatifs de 0.720 à 0.854 et les prévisions pré opératoires confirmées par les constatations per opératoires sont excellentes 87.4% ($k=0.823$). La classification d'Hartofilakidis est un système parfaitement viable qui permet de prévoir un pré opératoire les lésions que l'on pourra constater en per opératoire.

Introduction

For the classification of congenital hip disease (CHD) in adults various systems have been proposed [1, 12, 16, 17]. The Hartofilakidis et al. classification system, initially published in 1988, relies on intraoperative findings to describe the hip pathology encountered during the operation [1, 2]. The major distinguishing feature is the description of acetabular deformity (Table 1). This classification system encompasses three types of deformity in the adult hip, i.e. dysplasia, low dislocation and high dislocation (Fig. 1). The demographics and clinical presentation vary between the various CHD types [13]. Patients with dysplasia become symptomatic later in adult life, the anatomical distortion of the femur and the acetabulum is milder, the operation is simpler and the results of total hip replacement (THR) are generally better compared to the more severe forms of the disease [14]. In 20–50% of adults with hip osteoarthritis the underlying problem is hip dysplasia [9].

C. K. Yiannakopoulos · G. C. Babis · G. Hartofilakidis
First Orthopaedic Department,
University of Athens Medical School,
Athens, Greece

T. Xenakis
Department of Orthopaedic Surgery, School of Medicine,
University of Ioannina,
Ioannina, Greece

T. Karachalios
Orthopaedic Department, School of Medicine,
Faculty of Health Sciences, University of Thessaly,
Papakyriazi 22,
Larissa 41222, Greece

C. K. Yiannakopoulos (✉)
Byzantiou 2, Nea Smyrni,
Athens 17121, Greece
e-mail: cky@ath.forthnet.gr

Table 1 The Hartofilakidis et al. system for classification of CHD in adults

Type	Description	Deficiencies of the acetabulum as verified during surgery
Dysplastic hip	The femoral head is contained within the original acetabulum despite the degree of subluxation	<ul style="list-style-type: none"> •Segmental deficiency of the superior wall •Secondary shallowness due to fossa-covering osteophyte
Low dislocation	The femoral head articulates with a false acetabulum that partially covers the true acetabulum to a varying degree	<ul style="list-style-type: none"> •Complete absence of the superior wall •Anterior and posterior segmental deficiency •Narrow opening and inadequate depth of the true acetabulum
High dislocation	The femoral head is completely out of the true acetabulum and migrated superiorly and posteriorly to a varying degree	<ul style="list-style-type: none"> •Segmental deficiency of the entire acetabulum with narrow opening. •Inadequate depth •Excessive anteversion •Abnormal distribution of bone stock, mainly located superoposteriorly in relation to the true acetabulum

Ideally, a classification system in addition to being reliable should validly predict the intraoperatively anticipated structural bone deformities or abnormalities and aid treatment planning. It should also include all types of deformity, be simple, easy to memorise and accurate. Reliability is essentially the extent of the agreement between repeated measurements, and validity is the extent to which a method of measurement provides a true assessment of that which it purports to measure [3, 5].

The purpose of this study was to examine the interobserver reliability and the validity of the Hartofilakidis et al. classification of CHD in adults comparing radiographic morphology with intraoperative findings.

Materials and methods

The anteroposterior pelvis radiographs of 102 patients with hip osteoarthritis secondary to CHD were examined. The radiographs were obtained from the senior author's (GH) database. The total number of hips examined was 158. The study was approved by the Institutional Review Board.

In all cases the morphology of the hip joint and the pathology of hip deformity were assessed and recorded at the time of surgery by the senior author (GH). There were

22 cases (13.9%) with dysplasia, 70 cases (44.3%) with low dislocation and 66 cases (41.2%) with high dislocation.

Each radiograph was independently assessed by three experienced senior hip surgeons from different Universities of Greece. The observers were not involved in the selection of the radiographs and had no knowledge of the name, age and sex of each patient. All observers received a detailed description and a diagrammatic explanation of the Hartofilakidis et al. classification system along with a CD containing all radiographs. Interobserver testing was carried out in a blinded fashion. These results were then compared with the intraoperative findings as recorded by the senior author by another examiner. Interobserver reliability was assessed by examining the agreement between the three raters. The validity of the classification system was assessed by examining the agreement between the assessment by each one of the three raters and the intraoperative findings, which were used as a reference standard (Table 2).

Sample size estimation

The intraclass correlation coefficient is often used as an index of reliability in a measurement study. In these studies, there are N observations made on each of K individuals. These

Fig. 1 The radiological appearance of the three types of CHD in adults: dysplasia (a), low dislocation (b) and high dislocation (c)

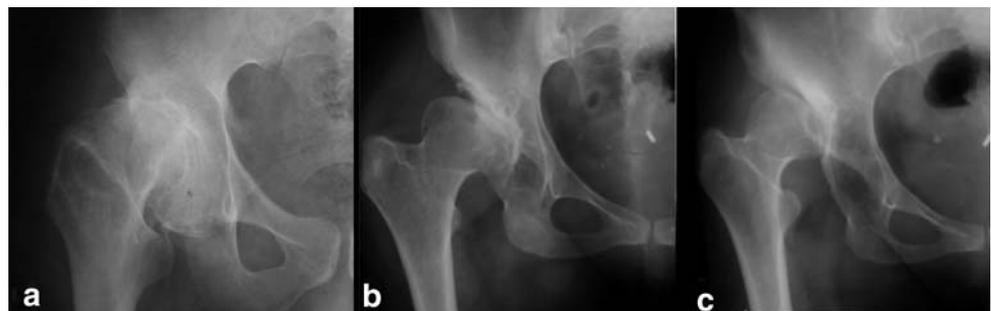


Table 2 Comparison of percent of agreement with the gold standard method (intraoperative classification) among the three observers

	Percent of agreement
Observer 1	85.7%
Observer 2	87.7%
Observer 3	88.7%
	$\chi^2=0.83, p=0.660$

individuals represent a factor observed at random. This design arises when K subjects are each rated by N raters.

The intraclass correlation coefficient may be thought of as the correlation between any two observations made on the same subject. When this correlation is high, the observations on a subject tend to match, and the measurement reliability is 'high'.

Sample size calculation was based on the primary outcome with the aim of showing a reliability that was at least substantial ($\kappa > 0.7$); the power was set to 90%, $\alpha = 0.05$, and $\beta = 0.10$. Using this approach, a sample size of $n = 3$ observers and $K = 20$ X-rays per observer was calculated for each group.

Statistical analysis

Assessment of interobserver consistency was accomplished using two parameters: the proportion of agreement and the kappa coefficient as proposed by Fleiss.

The observed proportion of agreement is the percentage of instances in which the observers agreed. The kappa coefficient involves adjustment of the observed proportion of agreement by correction for the proportion of agreement which arises due to chance.

Observer's agreement with the gold standard method was examined using Cohen's quadratic weighted kappa (K) coefficient. Interobserver agreement was assessed by calculating kappa coefficients for every possible pair of observers. Kappa is the chance-corrected proportional agreement, and possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete

disagreement). Interpretation of the data was performed according to Landis and Koch [11]. An agreement is graded as slight ($K = 0-0.2$), fair ($K = 0.21-0.40$), moderate ($K = 0.41-0.60$), substantial ($K = 0.61-0.80$) and almost perfect ($K = 0.81-1$). The observed proportion of agreement with the gold standard method among the observers was compared with the chi-square test. Pearson's chi-square test was also used to compare the distribution of the three types of hip deformity between the right and the left hip. The level of significance was $p < 0.05$. The statistical analysis was performed using SPSS version 13.00 statistical package.

Results

The agreement between the gold standard method and the three observers was excellent. The overall percentage of agreement of all three observers compared with the intraoperative judgment was 87.4% ($K = 0.823$, excellent agreement). The agreement for observer 1 was 85.7% ($K = 0.8$, excellent agreement, $p < 0.0005$), for observer 2 87.7% ($K = 0.831$, excellent agreement, $p < 0.0005$) and for observer 3 88.7% ($K = 0.839$, excellent agreement, $p < 0.0005$).

There was no statistically significant difference among the three observers concerning the percent of agreement with the gold standard method (Table 3).

The percentage of agreement between observer 1 and observer 2 was 80.3% ($K = 0.728$, substantial agreement), between observer 1 and observer 3 it was 87.2% ($K = 0.816$, excellent agreement) and between observer 2 and 3 it was 81.7% ($K = 0.740$, substantial agreement) (Table 4).

Discussion

Several classification systems have been used to describe the different types of CHD in adults [1-4]. However, the reliability of those systems has not explicitly been the focus of attention and their validity has not been, to our knowledge, been reported.

This study evaluated the interobserver reliability of the Hartofilakidis et al. classification system using preoperative

Table 3 The interobserver agreement between the three observers 1, 2 and 3. K kappa measure of agreement, SE standard error, Sig statistical significance, % agr % agreement

	Observer 2				Observer 3			
	K	SE	Sig	% agr	K	SE	Sig	% agr
Observer 1	0.728	0.038	<0.0005	80.3%	0.816	0.034	<0.0005	87.2%
Observer 2	–	–	–	–	0.740	0.037	<0.0005	81.7%

Table 4 The intraobserver agreement between the three observers 1, 2 and 3 for the left, the right and both hips. *K* kappa measure of agreement, *SE* standard error, *Sig* statistical significance, % agr % agreement

	Observer 2				Observer 3			
	Left hip				Left hip			
	K	SE	Sig	% agr	K	SE	Sig	% agr
Observer 1	0.721	0.053	<0.0005	80.4%	0.854	0.044	<0.0005	90.2%
Observer 2	–	–	–	–	0.722	0.054	<0.0005	80.4%
	Right hip				Right hip			
	K	SE	Sig	% agr	K	SE	Sig	% agr
Observer 1	0.720	0.056	<0.0005	80.2%	0.770	0.052	<0.0005	84.1%
Observer 2	–	–	–	–	0.747	0.053	<0.0005	82.1%
	Left and right hips				Left and right hips			
	K	SE	Sig	% agr	K	SE	Sig	% agr
Observer 1	0.728	0.038	<0.0005	80.3%	0.816	0.034	<0.0005	87.2%
Observer 2	–	–	–	–	0.740	0.037	<0.0005	81.7%

anteroposterior pelvis radiographs and examined its validity using the intraoperative findings as the gold standard method.

Reliability of a classification system depends on the consistency of measurements or observations and has to do with the quality of measurement or observation. It also describes the extent of the agreement between repeated measurements. A reliable classification system classifies a disease or fracture consistently but it does not necessarily reveal what in reality is happening. A valid classification system reveals the true underlying pathology of the disease or fracture [3, 5].

Validity is the best approximation to the “truth” providing a true assessment of that which it purports to measure or describe.

Reliability and validity are not independent but they are related to each other. A method may be reliable if it measures something consistently but is valid only if the result of measurement approximates the true value. Validity implies reliability but not vice versa. Reliability is a

necessary but not sufficient condition for validity. A classification system may be reliable but not valid, but it cannot be valid without being reliable [3].

Not every commonly used classification system in orthopaedics is reliable or reproducible. Substantial variation within the ratings of several radiograph reviewers may be noted. The accuracy of any classification system, which can be considered as a measuring instrument, is described estimating the reliability and the validity of the data. The quality of a measurement or observation can be described with the determination of reliability and validity. The more valid and reproducible a classification system is the better are the communication between clinicians and researchers, the treatment planning and the evaluation of the results of any given therapeutic intervention.

Preoperative radiographs are used as a projection of the pathology or the expected bone deficiencies encountered during hip surgery. The surgeon should be able to recognise on preoperative radiographs the anatomical abnormalities

Fig. 2 Anteroposterior radiograph and 3-D CT scan of a 34-year-old woman with CHD of the right hip. On the preoperative radiograph (a) raters 1 and 2 classified the hip as dysplastic and rater 3 as low dislocation. The 3-D CT scan (b) proved that the hip was dysplastic. Diagnosis was confirmed during total hip arthroplasty

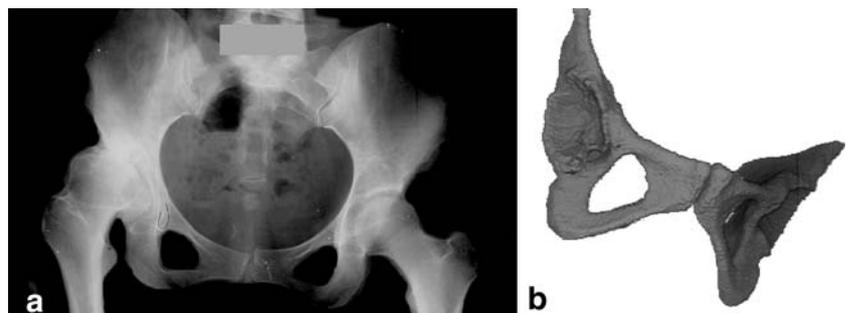




Fig. 3 Anteroposterior radiograph and 3-D CT scan of a 44-year-old woman with bilateral CHD. On the preoperative pelvis radiograph (a) the right hip was classified as high dislocation by all three raters. For the classification of the left hip there was disagreement. Raters 2 and 3

classified the hip as high dislocation and rater 1 as low dislocation. The 3-D CT scan (b, c) proved that in both hips the type of dislocation was high and the diagnosis was confirmed during total hip arthroplasty

which may be encountered during the operation. This facilitates preoperative planning and assists in effectively dealing with the technical difficulties of the operation.

In our study there was significant agreement between the preoperative rater's assessments (reliability) and the intraoperative assessment (validity) with weighted kappa values of >0.75 . These results may be partially due to the fact that the reviewers, though from different institutions, were familiar with the Hartofilakidis et al. classification having performed many THRs for CHD. This may be considered as a limitation of the study; however, we assumed that the more experienced the observers the better would be the understanding of the definitions of a classification system and the prediction of the expected bone deficiencies found intraoperatively.

In tentative cases the underlying anatomy of the hip joint can be evaluated with three-dimensional (3-D) computed tomography (CT). With exclusion of the femoral head from the final image the deficiencies of the acetabulum can be readily appraised (Figs. 2 and 3).

The inter- and intraobserver reliabilities of the Crowe and Hartofilakidis classifications were recently examined by Decking et al. [7]. In this paper the radiographs of 51 patients (62 hips with CHD) were included. According to those authors both systems can be recommended because of their high reliability; however, to our knowledge, validation of any CHD classification system has not yet been performed. This is the case because classifications such as the one of Crowe et al. [16] cannot be validated because they are based not on the pathology of the acetabulum or the femur but rather on arbitrary assumptions or measurements performed on radiographs. The Hartofilakidis et al. classification is based on the pathology of the acetabulum as determined intraoperatively.

This study evaluated the validity of the Hartofilakidis et al. classification using as a gold standard the intraoperative findings. There was a high correlation between the preoperative radiographs and the intraoperative findings among all participants in this study. The intraoperative

findings in contrast to the radiographic findings are considered soft data [13]. Observer bias to the validation process has been kept to a minimum since all intraoperative observations were made and recorded by the senior author.

The validity of several classification systems in revision total hip and knee arthroplasty has been evaluated comparing assessment of preoperative radiographs and intraoperative findings [10, 18].

A classification is purported to measure some future performance. The success of THR in CHD, as judged by the rate of revision surgery, depends on the severity of the anatomical distortion of the acetabulum which is explicitly described by the Hartofilakidis et al. classification. Survival of the THR in patients with CHD depends on the type of the acetabular deformity [8, 11, 14, 15], acetabular component loosening being the weak link.

The Hartofilakidis et al. classification system has been shown to be reliable as well as valid in predicting the acetabular abnormalities encountered during THR. It provides a reliable estimate of the acetabular bone loss in adults with hip osteoarthritis secondary to CHD. With this classification the structural alterations of the acetabulum could be reliably predicted using standard anteroposterior pelvis radiographs.

Knowledge of the structural deformities of the acetabulum in CHD in adults and of the segmental roof or wall defects anticipated during surgery, as predicted by the Hartofilakidis classification, improves preoperative planning and may improve cup placement, anticipating an increase in the endurance and longevity of the component.

The reliability and validity of the Hartofilakidis et al. classification system relates to its ability to consistently predict the pathoanatomical changes of the acetabulum and thus to apprise the surgeon of the different defect types anticipated during surgery.

Acknowledgement We wish to thank Antonis Galanos, M.Sc. for providing help with the statistical analysis.

References

1. 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
2. Chougle A, Hemmady MV, Hodgkinson JP (2006) Long-term survival of the acetabular component after total hip arthroplasty with cement in patients with developmental dysplasia of the hip. *J Bone Joint Surg Am* 88:71–79
3. 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
4. Decking R, Brunner A, Decking J, Puhl W, Gunther KP (2006) Reliability of the Crowe und Hartofilakidis classifications used in the assessment of the adult dysplastic hip. *Skeletal Radiol* 35:282–287
5. Eftekhari NS (1978) Principles of total hip arthroplasty. Mosby, St. Louis
6. Garbuz DS, Masri BA, Esdaile J, Duncan CP (2002) Classification systems in orthopaedics. *J Am Acad Orthop Surg* 10:290–297
7. Gozzard C, Blom A, Taylor A, Smith E, Learmonth I (2003) A comparison of the reliability and validity of bone stock loss classification systems used for revision hip surgery. *J Arthroplasty* 18:638–642
8. Greenfield ML, Kuhn JE, Wojtys EM (1998) A statistics primer. Validity and reliability. *Am J Sports Med* 26:483–485
9. 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
10. Hartofilakidis G, Stamos K, Karachalios T, Ioannidis TT, Zacharakis N (1996) Congenital hip disease in adults. Classification of acetabular deficiencies and operative treatment with acetabuloplasty combined with total hip arthroplasty. *J Bone Joint Surg Am* 78:683–692
11. Hartofilakidis G, Karachalios T, Stamos KG (2000) Epidemiology, demographics, and natural history of congenital hip disease in adults. *Orthopedics* 23:823–827
12. Hartofilakidis G, Karachalios T (2004) Total hip arthroplasty for congenital hip disease. *J Bone Joint Surg Am* 86:242–250
13. Kerboul M, Mathieu M, Sauzieres P (1987) Total hip replacement for congenital dislocation of the hip. In: Postel M, Kerboul M, Evrard J, Courpied JP (eds) Total hip replacement. Springer, New York, pp 51–66
14. Kim YH, Kim JS (2005) Total hip arthroplasty in adult patients who had developmental dysplasia of the hip. *J Arthroplasty* 20:1029–1036
15. Kocher MS, Zurakowski D (2004) Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Joint Surg Am* 86:607–620
16. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
17. Saleh KJ, Holtzman J, Gafni A, Saleh L, Davis A, Resig S, Gross AE (2001) Reliability and intraoperative validity of preoperative assessment of standardized plain radiographs in predicting bone loss at revision hip surgery. *J Bone Joint Surg Am* 83:1040–1046
18. Weinstein SL (1987) Natural history of congenital hip dislocation (CDH) and hip dysplasia. *Clin Orthop* 225:62–76