Original article

Influence of experience on intra- and inter-observer reproducibility of the Crowe, Hartofilakidis and modified Cochin classifications

A. Clavé [a,b,c,∗], L. Tristan [a,b], A. Desseaux [a,b], F. Gaucher [d], C. Lefèvre [a,b,c], E. Stindel [a,b,c]

[a] Université de Bretagne Occidentale, faculté de médecine, 22, avenue Camille-Desmoulins, 29200 Brest, France
[b] Service de chirurgie orthopédique et traumatologique de la Cavale-Blanche, boulevard Tanguy-Prigent, 29609 Brest cedex, France
[c] LaTIM, Inserm UMR 1101, 2, avenue Foch, 29609 Brest cedex, France
[d] Centre hospitalier de Cornouaille, 14, avenue Thépot, BP 1757, 29107 Quimper cedex, France

A B S T R A C T

Introduction: A previous study demonstrated that the reproducibility of the Crowe (Cr), the Hartofilakidis (Ha) and the modified Cochin (Co) classifications were comparable. However, there were differences with a trend that suggested the influence of experience. Therefore, we performed a prospective study to investigate whether experience influenced the reproducibility of the commonly used developmental dysplasia of the hip (DDH) classifications.
Hypothesis: The hypothesis was that the intra- and inter-observer reproducibility scores would be higher in the senior group than the junior group, and particularly for the modified Co classification.
Methods: Four seniors and four residents classified 104 A/P pelvic radiographs (200 hips) two times using the Cr, Ha and Co classification systems.
Results: For intra-observer reproducibility, the average weighted concordance coefficients [95% confidence intervals] were for the senior and the junior groups: 92.2 [88.6–95.7] and 92.6 [87.9–97.2] for Cr, 92.1 [88.7–94.6] and 92.0 [87.7–96.3] for Ha, 94.2 [91.8–96.6] and 94.1 [91.5–96.6] for Co. The average weighted Kappa (95% confidence intervals) were 0.8 [0.71–0.88] and 0.79 [0.68–0.89] for Cr, 0.77 [0.74–0.81] and 0.75 [0.62–0.88] for Ha, 0.82 [0.76–0.89] and 0.80 [0.74–0.87] for Co. The junior inter-observer reproducibility multi-rater Kappa (list A:list B) were 0.57:0.50 (Cr), 0.47:0.53 (Ha), 0.42:0.42 (Co). Senior multi-rater Kappa were 0.53:0.49 (Cr), 0.40:0.34 (Ha), 0.40:0.43 (Co).
Conclusions: Contrary to our hypothesis, the experience of the observer did not affect the intra- and inter-observer reproducibility of the three classification systems.

© 2016 Published by Elsevier Masson SAS.

1. Introduction

The incidence of developmental dysplasia of the hip (DDH) in adulthood is declining thanks to improvements in its prevention and its treatment during childhood [1]. However, treatments remain complex and rely on difficult surgical procedures mainly due to major anatomical bone as well as muscle and tendon changes [2–4]. Preoperative evaluation is one of the key components that help in defining the surgical strategy and may possibly predict the results [5].

Currently this evaluation uses radiographic classification systems, whether objectively based on the measurement of a radiographic index, or subjectively based on the assessment of descriptive anatomical elements [6]. Crowe (Cr; objective) [3] and Hartofilakidis (Ha; subjective) [4] classifications are the most widely used in the English-speaking world [7] while the SoFCOT (French Society of Orthopaedic and Trauma Surgery)-modified Cochin classification (Co; subjective), resulting from the work of Marcel Kerboull, is the reference used in France [2,6,8,9].

A previous study in 2014 [10] demonstrated that the reproducibility of the Co classification was comparable to the Cr and Ha classifications. However, while all observers of this study were senior French surgeons with a preference for the use of the Co classification, there were differences with a trend suggestive of influence of experience. The influence of experience has already been well demonstrated for several other orthopaedic radiological classifications [11,12]; however, it has been studied only once for Cr and Ha classifications [7] and its impact on the Co classification is unknown. One might expect that due to its subjective nature and its greater complexity (5 classes), the reproducibility of the Co

∗ Corresponding author. Service de chirurgie orthopédique et traumatologique de la Cavale-Blanche, boulevard Tanguy-Prigent, 29609 Brest cedex, France.
Tel.: +33 2 98 34 72 73; fax: +33 2 98 34 78 13.
E-mail address: arnaud.clave@orange.fr (A. Clavé).

classification is more prone to be influenced by the experience of the observer than the other two classifications.

The main objective of this study was to investigate whether the experience of the observer had an influence on the reproducibility of the commonly used DDH classifications. The hypothesis was that the intra- and inter-observer reproducibility scores would be higher in the senior group than the Junior group, and particularly for the modified Co classification.

2. Materials and methods

2.1. Patients

An independent observer selected a homogeneous series of analysable anterioposterior (A/P) pelvic views after screening the radiography databases from five orthopaedic centres specialising in DDH treatment. The eligible radiographs were from adult patients and based on the following inclusion and exclusion criteria:

- inclusion criteria:
  - A/P pelvis radiographs of a standing adult presenting signs of DDH on at least one hip as defined by the Co classification;
- exclusion criteria:
  - radiographs presenting no sign of DDH,
  - radiographs that did not show the entire pelvis (anteriosuperior iliac spine to the ischium),
  - radiographs not taken in full frontal view (defined as asymmetry of the iliac crests and obturator foramen and/or coccyx projection that was not centred relative to the pubic symphysis),
  - hips operated on with hip replacement or arthrodesis.

To reduce any memorisation of the results and recall bias, the radiographs were anonymised and the sequence in which they were analysed was randomly set by making up two reading lists (list A and list B).

2.2. Methods

Four experienced observers (orthopaedic consultants or professors; S1 to S4), and four novice observers (orthopaedic residents; J1 to J4) were divided into a senior group and a junior group. Each
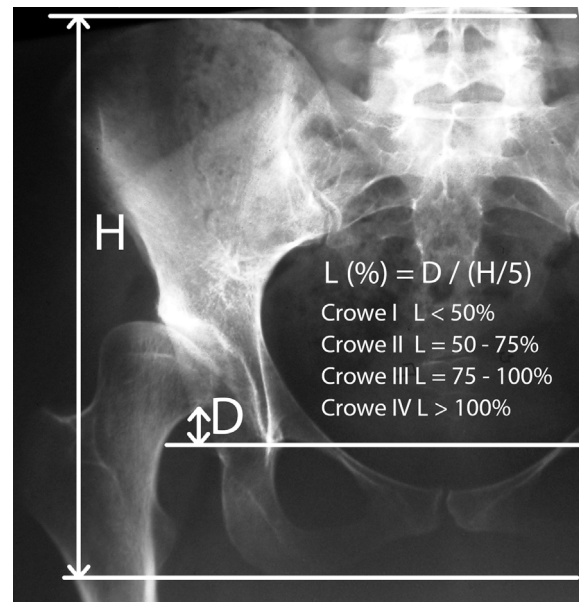

Fig. 1. Crowe Classification (Cr) [3] relies on the assumption that the normal ratio of the diameter of the femoral head to the height of the pelvis is 1:5 and that any proximal migration of the femoral head can be expressed either as a percentage of the height of the pelvis or of the height of the femoral head-neck junction. $L(\%) = D/(H/5)$. D: tear drop head-neck distance; H: height of the pelvis. Crowe I: $L < 50\%$; Crowe II: $L = 50–70\%$; Crowe III: $L = 75–100\%$; Crowe IV: $L > 100\%$.

participant was given two computer files two months apart including:

- the anonymised list of radiographs to analyse;
- a digitised version of the radiographs (list A in the first batch, list B in the second);
- a sheet used for the standardised recording of the grades;
- a reminder of the definitions of the three classifications used (Cr, Ha and Co) to avoid any biases related to free interpretation (Figs. 1–3 and Table 1).

Each analysable hip of each radiograph was classified according to the three classification systems adopted (Cr, Ha and Co) and the results were collected and frozen by an independent observer
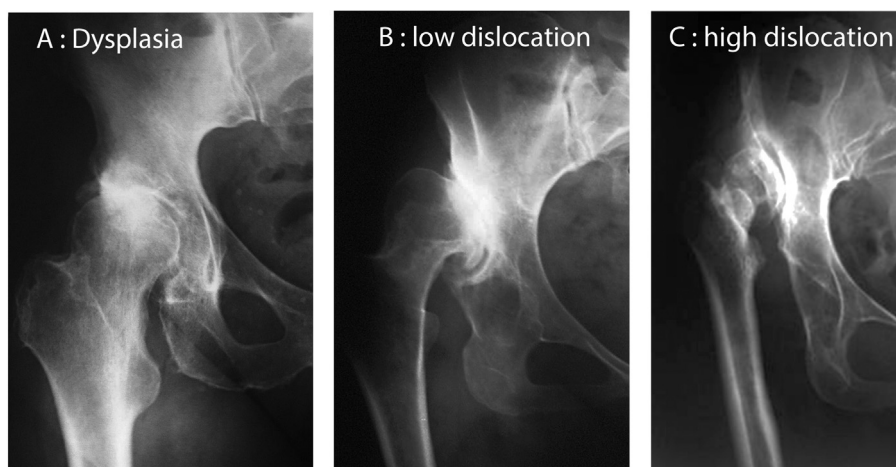
Fig. 2. Hartofilakidis Classification (Ha) [4] recognises three types of congenital hip disease: dysplasia and low and high dislocation. Dysplasia: the femoral head is contained within the original acetabulum despite the degree of subluation. Low dislocation: the femoral head articulates with a false acetabulum, which partially covers the true acetabulum to a varying degree. High dislocation: the femoral head is completely out of the true acetabulum and migrated superiorly and posteriorly to a varying degree.