Radiographic assessment in Perthes disease

Hip development and evaluation of prognostic factors

Stefan Huhnstock

Thesis for the degree of philosophicae doctor (PhD)

Faculty of Medicine, University of Oslo

Institute of Clinical Medicine

Division of Orthopedic Surgery, Oslo University Hospital

2018
Es bleibt hier viel zu erforschen

Georg Perthes’ final remark in his last publication (1924)
TABLE OF CONTENTS

1 Acknowledgment .................................................................................................................. 6

2 Abbreviations ....................................................................................................................... 8

3 List of papers ......................................................................................................................... 9

4 Introduction .......................................................................................................................... 10

4.1 What is Perthes disease? .................................................................................................. 10

4.2 Normal development of the child’s hip ............................................................................ 12

4.2.1 Acetabulum ................................................................................................................ 12

4.2.2 Proximal femur ......................................................................................................... 14

4.2.3 Arterial blood supply ............................................................................................... 16

4.3 Pathologic hip development in Perthes disease ............................................................. 19

4.3.1 Femoral head deformity ......................................................................................... 19

4.3.2 Growth disturbance of the proximal femur ............................................................. 22

4.3.3 Development of acetabular dysplasia ................................................................... 23

4.4 Imaging modalities in Perthes disease ......................................................................... 25

4.4.1 Conventional Radiography .................................................................................... 25

4.4.2 Other imaging modalities ....................................................................................... 26

4.5 Prognostic factors in Perthes disease ............................................................................ 27

4.5.1 Gender and Age ...................................................................................................... 27

4.5.2 Bilateral Perthes disease ....................................................................................... 28

4.5.3 Radiographic prognostic factors .......................................................................... 28

4.5.4 Long-term prognosis ............................................................................................. 30

5 Aims of study ....................................................................................................................... 32

6 Summary of papers I - IV .................................................................................................. 33

6.1 Paper I ............................................................................................................................. 33
6.2 Paper II .................................................................................................................................34
6.3 Paper III .................................................................................................................................35
6.4 Paper IV ................................................................................................................................36

7 General Discussion ...................................................................................................................38

7.1 Patients....................................................................................................................................38

7.1.1 Nationwide study on Perthes disease in Norway (Paper I- III) ........................................38
7.1.2 Retrospective review of hospital radiographic archive (Paper IV) ....................................39

7.2 Methods .................................................................................................................................40

7.2.1 Radiographic imaging .........................................................................................................40
7.2.2 Statistical analysis ...............................................................................................................51

7.3 Results ....................................................................................................................................55

7.3.1 Acetabular changes during the course of Perthes disease ..................................................55
7.3.2 Interobserver agreement and reliability ..............................................................................56
7.3.3 Prognostic factors for radiographic outcome ......................................................................62

8 Conclusions .............................................................................................................................66

9 Future Research .........................................................................................................................68

10 References ...............................................................................................................................69
1 Acknowledgment

In the final stage of my residency at the department of orthopedic surgery, Oslo University Hospital, my wonderful colleagues Vera Halvorsen and Ola Wiig introduced me to the field of pediatric orthopedic surgery. With their extensive expertise and winning collegiality, they inspired me to continue with a subspecialty training within pediatric orthopedics. At the same time Ola Wiig presented me with the idea of a research project in close conjunction to the Norwegian nationwide Perthes study. With formalization of this PhD project in 2010, a wonderful and exciting scientific journey began.

The Norwegian nationwide study on Perthes disease was conducted under the patronage of the Norwegian Pediatric Orthopedic Society. I want to thank all participating patients and all colleagues of the 28 hospitals throughout Norway, who collected and reported all necessary data. Without their contribution and commitment the study would not have been possible.

I would like to thank the former head of the department of orthopedic surgery, Oslo University Hospital Ullevål, Lars Engebretsen, for giving me the opportunity to join the department and the assigned PhD program. Further thanks to Rolf Riise, the head of the newly formed Orthopedic Division at the Oslo University Hospital at which I conducted and finalized my PhD thesis.

This thesis was further supported by the Sophies Minde Research Foundation and a Smith & Nephew research grant.

To carry out a research project, which is embedded in a university educational program, requires a huge amount of academic freedom, uncoupled from clinical routines. I am much obliged to my dearest colleagues of the pediatric orthopedic department, Rikshospitalet, Andreas Knaus, Per Reidar Høiness, Anders Wensaas, Anne Berg Breen and Ivan Hvid, for
supporting my work and stepping in for me whenever I was in need. I would like to express my sincere gratitude to the head of our department, Joachim Horn, for his continuous and boundless support throughout my project.

Numerous authors have contributed substantially to the studies included in this thesis and my special thanks go to Else Merckoll (Department of Radiology, Oslo University Hospital), Are Hugo Pripp (Department of Biostatistics, Epidemiology and Health, Oslo University Hospital), Svein Svenningsen (Department of Orthopedic Surgery, Sørlandet Hospital, Arendal) and Anthony Catterall (Royal National Orthopedic Hospital, London, England)

I want to deeply thank my co-supervisor Terje Terjesen who initiated the nationwide study. With his great scientific experience and meticulous editorial approach, he has been working diligently, helping me to improve each of the articles as well as the thesis.

I consider myself very fortunate to have Ola Wiig as my clinical mentor and main supervisor for this PhD thesis. With skillful guidance, enthusiasm and scientific scrutiny, he has let me grow and unfold into the world of science and research. This work would not have been possible if it was not for his support, encouragement, inspiration and last but not least his patience with me.

I would like to thank my parents, Renate and Jürgen Huhnstock, as well as my sister Gritt and her family for their ever-lasting support. They are the fundament and pillar I always can rely on.

Finally, I am deeply indebted to my wife and best friend Katrin Schauer, and our wonderful daughter Paula for their continuous love and patience. Their laughter, warmth and affection gave me the strength and energy to succeed with this project.
## 2 Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>2 dimensional</td>
</tr>
<tr>
<td>ADR</td>
<td>Acetabular- Depth- Width Ratio</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AP</td>
<td>Antero-posterior</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>DDH</td>
<td>Developmental dysplasia of the hip</td>
</tr>
<tr>
<td>NGP</td>
<td>Femoral Neck growth plate</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>ISS</td>
<td>Ischial Spine Sign</td>
</tr>
<tr>
<td>LAI</td>
<td>Lateral acetabular inclination</td>
</tr>
<tr>
<td>LGP</td>
<td>Longitudinal growth plate</td>
</tr>
<tr>
<td>LLA</td>
<td>Lateral lip of the acetabulum</td>
</tr>
<tr>
<td>MED</td>
<td>Multiple epiphyseal dysplasia</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>N</td>
<td>Number of patients</td>
</tr>
<tr>
<td>Obs</td>
<td>Observers</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>TGP</td>
<td>Trochanteric growth plate</td>
</tr>
</tbody>
</table>
3 List of papers


4 Introduction

4.1 What is Perthes disease?

More than 100 years ago, 3 surgeons independently described a distinct hip affection occurring in children and adolescents (Calvé 1910, Legg 1910, Perthes 1910). Over the past decades this condition has been termed either referring to the above authors (Legg-Calvé-Perthes disease, Legg-Perthes disease, Calvé-Legg-Perthes disease, Perthes disease) or describing the nature of this condition such as arthritis deformans juvenilis, osteochondritis deformans coxae juvenilis, coxa plana among others.

Perthes disease is a hip condition, which impairs the normal growth and development of the affected hip. The femoral head initially suffers from a partial or complete necrosis, which results in softening of the epiphysis and later a gradually deformity. After fragmentation of the epiphysis, development of the new bone formation can be observed and the femoral head eventually heals with a varying degree of deformity. The final shape of the hip can range from a nearly normal joint configuration to an extensive deformation with flattening and subluxation of the femoral head, broadening of the femoral neck and a dysplastic acetabulum.

Perthes disease has been described in children between 2 and 12 years of age, but it is most common between 4 to 8 years. It is 4 times more commonly seen in boys than in girls (Loder and Skopelja 2011). In the majority of the cases one hip is affected, but bilateral involvement has been reported in 8%-25% of children (Guille et al. 2002).

A critical distinction has to be made between bilateral Perthes disease and other conditions affecting the femoral head epiphysis bilaterally. Meyers dysplasia (1964) occurs in younger children (below the age of 4) (Khermosh and Wientroub 1991), but its radiographic appearance differs from Perthes disease (Harel et al. 1999, Rowe et al. 2005), with no
fragmentation, condensation or subluxation of the epiphysis (Meyer 1964). The prognosis is uniformly good and the hips recover gradually within 2-4 years to a normal or almost normal shape of the femoral head. Another bilateral femoral head affection that can resemble Perthes disease is the multiple epiphyseal dysplasia (Fairbank 1947). It is a form of skeletal dysplasia, which may cause disorganized enchondral ossification of the epiphyses of long bones and vertebrae (Briggs et al. 1993). Unlike Perthes disease, there are more pairs of epiphyses affected, without metaphyseal involvement (Anthony et al. 2015). Weight-bearing joints are predominately involved (Unger et al. 2008), and the articular cartilage degenerates rapidly, leading to an early osteoarthritis (Treble et al. 1990).

The initial clinical presentation of Perthes disease can differ widely, but it is not uncommon that the examination of the hip appears to be almost normal early on. However, reduced range of motion, especially for abduction and internal rotation, is one of the earliest clinical signs to be detected. As the disease advances, clinical symptoms may aggravate due to alteration of the proximal femoral geometry. Children with a deformed hip joint experience impaired hip joint function and the residual deformity predisposes for early onset osteoarthritis. In order to understand the complexity of the impaired hip development in Perthes disease, we have to have a better understanding of the normal development of a child’s hip.
4.2 Normal development of the child’s hip

4.2.1 Acetabulum

Ponseti (1978) characterized the acetabular anatomy and development in the postnatal period with special regard to the triradiate cartilage. The cartilaginous acetabulum is composed of a cup- shaped acetabular cartilage laterally and a Y-shaped triradiate cartilage medially.

![Lateral view](image1.png) ![Medial view](image2.png)

Figure 1. Normal acetabular cartilage complex of a 1-day-old child. Bony parts of the ilium, ischium and pubis have been removed. The lateral view shows a cup-shaped acetabulum and the medial view shows the 3 flanges of the triradiate cartilage. Reprinted from “Growth and development of the acetabulum in the normal child. Anatomical, histological, and roentgenographic studies.” by I. V. Ponseti, 1978, J Bone Joint Surg Am; 60 (5): 575-85.

The triradiate cartilage separates the bony parts of the ilium, ischium and pubis in the area of the acetabulum. The cup-shaped acetabular cartilage forms the outer two-thirds of the acetabular cavity, while parts of the ilium and ischium together with the triradiate cartilage form the non-articular medial wall of the acetabulum. The pubic bone is separated from the cavity by a thick cartilage. The acetabular cartilage complex is composed mostly of cellular hyaline cartilage and is covered by growth-plate cartilage where it adjoins with pelvic bones and articular cartilage where it articulates with the femoral head. The fibro-cartilaginous labrum forms the peripheral margin of the acetabular cartilage, increasing the relative depth of the socket.

The acetabular cartilage is an epiphysis and its development shows many similarities to the iliac crest and the epiphysis of long bones with the appearance of 3 secondary ossification centres until maturation. The os acetabuli is the functional epiphysis of the os pubis and it accounts for substantial development of the anterior acetabular wall. The iliac ossification centre is situated superiorly and contributes to shape of the lateral acetabular margin. The smallest of the 3 ossification centres is attributed to the os ischium and forms the posterior wall of the socket. All ossification centres appear by 8-9 years of age (Wiberg 1953, Ponseti 1978, Fabricant et al. 2013, Morris et al. 2015), and fuse in most cases until skeletal maturity.

Growth of acetabular height and width depends on interstitial growth of the triradiate cartilage. Growth in depth and the construction of the final acetabular shape is, however, heavily depended on the interaction with a spherical femoral head (Harrison 1961) as it functions as a template about which the acetabulum forms. It is known that severe Perthes disease may lead to a femoral head that is aspherical and flattened. This deformity has a
great impact on the interplay between acetabulum and femoral head in the growing hip as it may imbalance the conjoint development.

4.2.2 Proximal femur

The femoral head consists of an epiphysis, which is separated from the bony part of the femoral neck by the longitudinal growth plate (LGP) (Siffert 1981). The femoral neck is covered by a growth zone along the lateral border, termed femoral neck growth plate (NGP). A third growth plate is located at the basis of the greater trochanter, termed trochanteric growth plate (TGP). While the head of the femur and the greater trochanter enlarge by appositional growth, the 3 growth plates work simultaneously to support the longitudinal growth and to develop the shape of the proximal femur (Siffert 1981).

Figure 2. Structure of the proximal femur growth complex. It consists of 3 growth plates, which are namely the longitudinal growth plate (LGP), the femoral neck growth plate (NGP) and the trochanteric growth plate (TGP).

The LGP is anatomically located within the femoral head during infancy and contributes initially to its spherical development. As the neck elongates, the LGP moves relatively
laterally and is finally located at the femoral head-neck junction. The LGP grows proximally and medially, contributing mainly to the longitudinal growth of the femur and the neck. The TGP contributes mainly to longitudinal growth of the femur and the lateral width of the neck. Additionally, it points the growth of the greater trochanter in a proximal-lateral direction thereby giving the trochanter its characteristic shape. The NGP, a relatively small cartilage complex, accounts for growth of the femoral neck in width. Since there are no growth zones along the medial border of the neck, the growth of this lateral situated complex contributes essentially to varus and valgus angulation of the femoral head. The angular development is depended on the orientation of the respective growth plate areas. While TGP and NGP have growth vectors that are orientated laterally in the proximal femur, the LGP promotes growth in medial proximal direction relative to the femoral shaft. During infancy and early childhood, LGP is relatively horizontal and perpendicular to the long bone axis of the femur. As the growth rates of TGP and NGP increase towards adolescence, the LGP begins to tip medially, allowing for a more medially directed growth to balance proximal growth of the femur. A disturbance in any one of these growth plates imbalances their dynamic relationship, which may lead to angular deformities and disproportions of the proximal femur as seen in Perthes disease (for details see chapter 4.3.2).

Just as a located femoral head is necessary for acetabular development, the principle of hip joint containment accounts likewise for a normal femoral development (Siffert 1981). Contact pressures exerted on the femoral head cartilage by the acetabulum regulate appositional growth of the femur. Increasing pressures inhibit growth, compelling the femur to adapt to the spherical form of the acetabular template. It is well accepted that proximal femur and acetabular development are inseparably linked to achieve a congruent joint.
4.2.3 Arterial blood supply

The proximal femoral arterial supply consists of 3 entities: (1) an extracapsular arterial ring, (2) intracapsular ascending neck arteries and (3) an intracapsular subsynovial ring (Chung 1976). The extracapsular arterial ring is located at the base of the femoral neck and is formed by the branches of the medial and lateral circumflex arteries. Thin ascending vessels derive from that ring and pierce the hip capsule, further ascending in a subsynovial, intraarticular location along the femoral neck. The ascending vessels can be grouped according to their course: lateral, posterior, medial and anterior. Some of the branches from the ascending arteries pierce the neck of the femur to supply the metaphysis and parts of the greater trochanter. In large part the ascending vessels unite more proximally at the head-neck junction to form the intracapsular subsynovial ring anastomosis. From here, epiphyseal and metaphyseal branches arise. The epiphyseal branches cross the physeal plate bypassing the perichondral ring superficially and then enter the cartilage of the developing femoral head. The metaphyseal branches pierce the neck and run distally.

Figure 3. Posterior view of a right femur of 6.5-year-old girl (perfused with barium sulphate). The lateral ascending cervical artery divides to form metaphyseal branches, which pass into the neck (A) and superficial epiphyseal branches (B) to the ossification centre in the femoral head. Reprinted from “The arterial supply of the developing proximal end of the human femur.” by S. M. Chung, 1976, J Bone Joint Surg Am, 58 (7): 961-70. Copyright 2018 by Wolters Kluwer Health, Inc. Reprinted with permission.
Vascularization of the immature acetabulum involves 2 independent systems: the central axis and the peripheral ring (Damsin et al. 1992). The obturator artery emits a substantial branch, the acetabular artery, to the centre of the acetabulum, with nutrient vessels to the acetabular fossa (Itokazu et al. 1997), the triradiate cartilage and all 3 primary ossification centres (Damsin et al. 1992). To specify the blood vessel contributions to the outer acetabular margin, the acetabulum is divided in sectors as the face of a clock, seen from lateral. The upper dome (from 10 o’clock to 4 o’clock) is supplied by branches of the superior gluteal artery; at 4 o’clock to 8 o’clock, a posterior branch of the obturator artery supplies the inferior acetabular bone; at 8 o’clock to 10 o’clock an inferior branch of the inferior gluteal artery provides nutrient acetabular branches (Beck et al. 2003). A deep branch of the medial circumflex artery contributes to the arterial blood supply of the anterior wall of the acetabulum (Beck et al. 2003, Seeley et al. 2016).

The vascularization of the immature hip differs substantial to that of an adult and the vessel configuration of the proximal femur changes distinctly in the maturation process (Trueta 1957, Chung 1976). Three major differences illustrate the vulnerability of the femoral head blood supply and they may be of importance in the pathogenesis of Perthes disease:

(1) From birth until physeal plate closure, the growth plate is a vascular barrier and no vessel cross it. The ascending neck vessels run along the perimeter of the femoral neck and bypass the plate in the periphery to enter the femoral head (Crock 1980). The superficial course of the vessels makes the arterial blood supply vulnerable to pressure and mechanical exposure, which is one theory why Perthes only occurs in children.

(2) Each of the femoral neck arteries that penetrates into the femoral head, supplies a distinct anatomical area with arterial blood. The independent vascular territories of the epiphysis converge first in later growth to a larger anastomotic network. Occlusion of a specific ascending neck artery may cause necrosis of a vascular independent area, not being provided with a safeguard of an anastomotic network. The variable extent of femoral head necrosis as seen in Perthes disease may be a result of one or more vascular occlusions.

(3) The contributions of the medial and lateral circumflex artery to the arterial blood supply of the femoral head changes with maturation. The contribution of the medial circumflex artery reaches greater significance, since the number of arteries provided by the lateral circumflex artery decreases significantly with hip development (Chung 1976, Dewar et al. 2016). The transition to a more predominant single-stem arterial supply of the epiphysis increases the risk of an extensive necrosis in case of vascular obliteration, which may explain more pronounced changes as seen in older children with Perthes disease (Wiig et al. 2008).
4.3 Pathologic hip development in Perthes disease

Numerous etiologies have been proposed as a cause of this hip affection, including trauma, rickets, infections, congenital anomalies, rheumatic diseases and endocrine disorders. The theory that single or multiple events of blood supply interruptions to the femoral head may cause the deformity has been proposed by many authors (Jonsater 1953, Sanchis et al. 1973, Inoue et al. 1976). Clinical and experimental evidence supports the theory that disruption of the femoral perfusion is the key pathogenic event leading to characteristic changes seen in the course of the disease. At an early stage of the disease, angiography (Theron 1980, Atsumi et al. 2000), bone scintigraphy (Conway 1993), Perfusion MRI (Lamer et al. 2002) and biopsy studies (Jonsater 1953) showed clear evidence of a disruption of the blood supply to the femoral head consistent with ischemic necrosis.

4.3.1 Femoral head deformity

One of the greatest challenges to a better understanding of the pathogenesis is the limited access to clinical specimens. Per date there are 6 whole femoral heads and some clinical studies based on biopsies and necropsies in humans available to assess the histopathology and the major changes are summarized in Table 1 (Jonsater 1953, Ponseti 1956, Dolman and Bell 1973, McKibbin and Ralis 1974, Inoue et al. 1976, Catterall et al. 1982, Catterall et al. 1982, Ponseti et al. 1983).
Table 1. Summary of histopathological changes seen as a result of Perthes disease

<table>
<thead>
<tr>
<th>Anatomical structure</th>
<th>Histopathological changes in order of their appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Articular cartilage</strong></td>
<td>Increased thickness</td>
</tr>
<tr>
<td></td>
<td>Chondrocyte necrosis in the deep layer</td>
</tr>
<tr>
<td></td>
<td>Cessation of enchondral ossification</td>
</tr>
<tr>
<td></td>
<td>Separation of cartilage from subchondral bone</td>
</tr>
<tr>
<td></td>
<td>Vascular invasion</td>
</tr>
<tr>
<td></td>
<td>New accessory ossification</td>
</tr>
<tr>
<td><strong>Bony Epiphysis</strong></td>
<td>Necrosis of bone marrow and trabeculae</td>
</tr>
<tr>
<td></td>
<td>Fracture of trabeculae</td>
</tr>
<tr>
<td></td>
<td>Fibrovascular granulation tissue invasion</td>
</tr>
<tr>
<td></td>
<td>Osteoclastic resorption of necrotic bone</td>
</tr>
<tr>
<td></td>
<td>New bone formation with thickened trabeculae</td>
</tr>
<tr>
<td><strong>Physis</strong></td>
<td>Focal areas of growth cartilage extending into the metaphysis</td>
</tr>
<tr>
<td></td>
<td>Premature growth arrest in 30% of patients</td>
</tr>
</tbody>
</table>

The development of animal models such as the piglet model was of great importance to investigate the evolution of the hip after the disruption of the femoral head perfusion. Key findings were decreased mechanical properties of the femoral head after the ischemic event (Pringle et al. 2004, Koob et al. 2007). In the necrotic bone osteoblast activity was impaired, dysbalancing normal bone remodeling. An overload of osteoclast resorption led to
radiolucent areas in the epiphysis and contributed to a further weakening of the mechanical
stability of the infarcted head (Kim and Su 2002). Antiresorptive agents, like
bisphosphonates, has been shown to decrease the bony deformity in animal models,
supporting the theory that osteoclastic activity may be a contributing factor of femoral head
deformity in Perthes disease (Kim et al. 2005). The pathogenic concept of ischemia leading
to epiphyseal weakening - mechanical loading of a weak femoral head progresses
subsequently to deformity seems plausible. However, there is no data available in the
literature, defining the threshold of mechanical loading upon a weak femoral head.

Articular cartilage of the immature hip is thicker than that of the adult hip. Nutrition of the
outer layers of cartilage is provided by synovia while the deepest layers are dependent on
subchondral blood supply (Kim et al. 2001, Ytrehus et al. 2004). Subsequently, femoral head
ischemia initiates not only necrosis of the bony epiphysis but also damage to the deeper
layers of the cartilage surrounding the secondary center of ossification. The consequence is a
growth arrest of the secondary center (Kim et al. 2001).

Remodeling of the femoral head and subsequent healing depends on revascularization of
the femoral epiphysis. In a recent study by Kim et al. (2016), a novel method of serial
perfusion MRI was presented to elucidate the revascularization process in the necrotic
femoral heads of patients with Perthes disease. The authors observed that ingrowth of
vessels into the epiphysis varied from one patient to the other. Interestingly,
revascularization arose from the periphery and not centrally as reported in earlier studies
based on scintigraphy (Conway 1993, Tsao et al. 1997). Gradual restoration of femoral head
perfusion provides the revascularized regions with cells and soluble factors, enhancing
resorption of necrotic bone and new bone formation. The revascularization was first visible
at posterior, lateral and medial aspect, converging toward the central- anterior region over
time. It was observed that revascularization seemed not to proceed symmetrical, which potentially might contribute to asymmetrical growth and reconstruction of the femoral epiphysis.

4.3.2 Growth disturbance of the proximal femur

The altered growth of the proximal femoral in Perthes disease is a result of a combined disturbance of both the epiphyseal and the longitudinal growth plate in the metaphysis. Premature closure of the LGP occurred in 30% of patients with Perthes disease (Bowen et al. 1982, Sponseller et al. 1989) and it may be located either centrally or peripherally. In both cases the length of the femoral neck will be gradually reduced compared with the unaffected hip.

![Central arrest, Medial arrest, Lateral arrest](image)

Figure 5. Patterns of growth arrest of the longitudinal growth plate of the proximal femur.

In a central growth arrest, a short neck occurs without angular deformity of the neck. In decentralized growth arrest, the neck tends to grow with a deformity dependent to the location of the arrest (Siffert 1981), leading to either coxa vara or valga. Since trochanteric growth is not deprived in Perthes disease, this may result in relative overgrowth of the greater trochanter. The appositional growth along the femoral neck (NGP) is unaffected by
Perthes disease throughout the process and may yield a broad femoral neck. While the inner layers of the articular cartilage are dependent on subchondral vascularity, the outer layers derive nourishment from synovial fluid. The appositional growth of the superficial layers is not comprised by the ischemic event, thus they continue to grow, resulting in coxa magna. Enlargement of the femoral head leads to uncovering from the acetabulum. The uncovered lateral portion of the femoral head may be prone to unproportional growth, since growth regulating counter pressure of the acetabulum is missing. Hypertrophy on the lateral aspect prevents the femoral head to enter the acetabulum in abduction, and may force the femoral head to hinge on the acetabular rim. Excessive point pressure, exerted on a relative soft femoral head, compresses the matrix and may flatten the femoral head.

### 4.3.3 Development of acetabular dysplasia

Acetabular changes such as increased obliquity and loss of concavity have been noted since the first description of Perthes disease. Heyman and Herndon (1950) emphasized that dysplastic changes of the acetabulum played a crucial role in the evolution of Perthes disease and they defined it as 1 of 4 major radiographic criteria to describe the condition. Acetabular shallowness was later confirmed as a radiographic characteristic by several authors (Bellemans et al. 1979, Joseph 1989). Grzegorzekowski et al. (2006) found plastic deformation of the lateral upper margin of the acetabulum on AP pelvic radiographs. They described that the lateral margin was sloping upward in the majority of severe cases. Several authors found a thickening of the medial cartilage of the acetabulum that possibly was due to inflammatory processes and increased metabolism leading to femoral lateralization and a reduced relative depth of the acetabulum (Halkier 1956, Gershuni et al. 1981, Yngve and Roberts 1985). In addition, some authors highlighted that this hypertrophy occurred especially centered around the triradiate cartilage, dividing the acetabular concavity in 2
separate compartments (Yngve and Roberts 1985). It was hypothesized that deformed footprint of the acetabulum would form the femoral head, causing further flattening and deformity. Retroversion of the acetabulum has been reported in almost half of adult patients with a history of Perthes disease (Eijer et al. 2006, Ezoe et al. 2006). In a study that investigated retroversion in children with active Perthes disease, the authors found evidence that nearly 90% of the acetabula were retroverted (Larson et al. 2011).

Even though it is evident that Perthes disease not only affects the femoral head, little emphasis has been placed on research regarding the acetabular changes compared to the research efforts made, assessing femoral head changes. Most of the studies are of retrospective nature and investigated the acetabulum at a certain stage of the disease, without following up on the further development as the disease progressed. Our interest was to describe acetabular changes at different stages of Perthes disease and investigate how they are related to changes seen in the femoral head.

Radiographic measurements of the acetabulum are widely used in hip pathologies affecting foremost the acetabulum, such as in DDH. Acetabular measurements like the Sharp’s angle, Acetabular depth-length ratio (ADR) and lateral acetabular inclination have been validated thoroughly in these conditions. Since age at onset, degree of skeletally maturity and the nature of the deformity differ substantially from Perthes disease, conclusions regarding their validity cannot automatically be transferred. One objective was therefore to assess the interobserver agreement and reliability of measurements describing acetabular dysplasia in Perthes disease.
4.4 Imaging modalities in Perthes disease

4.4.1 Conventional Radiography

Radiography is an imaging technique that uses electromagnetic radiation (x-rays). A heterogeneous beam of x-rays is projected toward the field of interest and a certain amount of x-rays is absorbed depending on the particular density and composition of the structures. The X-rays that pass through the object are detected and they provide a superimposed 2D image of the internal structures.

The discovery of x-rays made it possible to describe the distinct hip changes found in Perthes disease and distinguish this condition from other hip pathologies (Calvé 1910, Legg 1910, Perthes 1910). As a 3-dimensional deformity is projected on a 2 dimensional image, several projections have to be obtained to capture most of the morphologic aspects. In Perthes disease, the AP pelvis and frog-leg lateral view of the hip have been considered as standard projections. The most commonly used classifications of the femoral head necrosis in Perthes disease are based on radiographic imaging (Catterall 1971, Stulberg et al. 1981, Herring et al. 2004). The advantages of radiography are the availability, the relative low costs and the fast recording time. In particular bony changes can be imaged in detail. However, the use of ionizing radiation (especially in a pediatric population) and its ineffectiveness to image soft tissue properly, delineates the limitations of conventional radiography. Furthermore, limited information can be gained of a 3 dimensional structure, which is projected 2-dimensionally. This poses obvious restrictions in the evaluation of complex deformities as discussed later (chapter 7.2.1).
4.4.2 Other imaging modalities

CT scan uses digital geometry processing to generate a 3-dimensional image of an object. It uses a large series of 2 dimensional radiographic images taken around a single axis of rotation (Herman and Herman 2009). CT scans can provide sliced images in various planes and reformatted 3D representations of structures, so that hip deformities can be depicted in detail. CT is regarded as a high-radiation diagnostic technique, which raises concerns of the use in the pediatric population. Hence, the use of CT in Perthes disease has been restricted to specific inquiries rather than being implemented as a standard diagnostic tool (Podeszwa and DeLaRocha 2013).

MRI is based on nuclear magnetic resonance. Hydrogen atoms, which exist in human fat and fluids, can absorb and emit radio frequency energy when placed in an external magnetic field. Since the first report of MRI in Perthes disease in 1984 (Scoles et al. 1984), the technique has been revolutionized, providing nowadays images of high anatomical precision. Numerous authors have investigated the value of this imaging technique in Perthes disease. Sanctis et al. (2011) concluded that MRI was advantageous in Perthes disease due to earlier detection of the femoral head necrosis and involvement of the physeal growth plate. Further, a prospective study (de Sanctis et al. 2000, de Sanctis and Rondinella 2000) investigated the prognostic value of femoral head necrosis, physeal and metaphyseal involvement obtained with MRI at time of diagnosis. The authors reported a significant association with the femoral head sphericity as measured by the Stulberg classification. Despite the early and accurate delineation of the morphologic changes, MRI has not yet been established as a standardized diagnostic tool in the evaluation of Perthes disease. At present, MRI is not equally available in all parts of the world and high costs may restrict the access for patient diagnosis. MRI is very time consuming, can be frightening because of noise
and no active movements are tolerated to avoid artifacts. The latter reasons pose challenges not that easy to overcome in a pediatric patient population. In a study investigating the feasibility of MRI imaging in Perthes patients, over 50% of the children needed either sedation of general anesthesia to conduct the examination (Sankar et al. 2014). This highlights that MRI is in the majority of the cases an invasive imaging method and therefore indications have to be considered with care.

### 4.5 Prognostic factors in Perthes disease

#### 4.5.1 Gender and Age

**Gender**

Several authors have reported that the final outcome in Perthes disease was worse in girls than in boys (Catterall 1971, Dickens and Menelaus 1978, Gershuni 1980, Mukherjee and Fabry 1990). Girls reach skeletal maturity earlier than boys and it was assumed that less time for remodeling of the femoral head would account for adverse outcome (Dickens and Menelaus 1978). However, other investigators could not find gender related differences for final outcome of Perthes (Guille et al. 1998, Wiig et al. 2008). Thus, there is no strong evidence to suggest that the prognosis is uniformly worse in girls than in boys.

**Age**

Numerous studies reported that age at the time of diagnosis is an important prognostic factor. The younger the child at onset, the better is the expected outcome. Children with onset before the age of 6 years were found to have significantly better outcome (Fabry et al. 2003, Rosenfeld et al. 2007, Wiig et al. 2008) than those presenting after the age of 8 years (Herring et al. 2004, Kamegaya et al. 2005). Perthes disease with onset in adolescence healed in most cases with poor result (Ippolito et al. 1985, Joseph et al. 2003).
4.5.2 Bilateral Perthes disease

Bilateral Perthes disease occurs less frequent than unilateral involvement and the number of reports on bilateral Perthes disease is limited. Most of the studies have a small number of patients and are of retrospective nature. Some authors investigated whether bilateral affection may influence the final radiographic outcome compared with unilateral affection and the results were remarkably divergent. Guille et al. (2002) reported that Perthes disease would run a milder course, with less deformity, whereas others suggested that bilateral affection would cause more pronounced hip deformity (Van den Bogaert et al. 1999), hence representing a more severe entity. The scope of our work was to describe the course of untreated bilateral Perthes hip disease and investigate the association to the residual hip deformity at a later stage.

4.5.3 Radiographic prognostic factors

Hips with Perthes disease develop with a distinct radiographic pattern (Waldenström 1922). Several authors recognized that the shape of the healed hip could vary, from being almost normal to incongruent and flat (Sundt 1921, Perthes 1924). It was noted that the hip deformity at skeletal maturity was associated with the extent of changes seen early in the course of the disease. The most important initial radiographic changes are: femoral head necrosis, subluxation of the femoral head, and accompanying acetabular dysplasia.

4.5.3.1 Femoral head necrosis

Numerous radiographic signs have been described reflecting the extent of femoral head necrosis and many of them have been associated with the final radiographic outcome. As an example, the extent of a subchondral fracture in an early phase was associated with the late radiological outcome (Waldenström 1922) and it was attributed a prognostic value (Caffey
1968, Catterall 1971, Salter and Thompson 1984, Wiig et al. 2004). Increased epiphyseal density (Bobechko and Harris 1960, Dolman and Bell 1973) and size of the epiphyseal nucleus (Caffey 1968), as well as the epiphyseal inclination angle (Catterall 1971) have been proposed to have significance for the radiographic development. The extent of femoral head necrosis in early fragmentation was characterized and linked to the final radiographic outcome by Catterall (Catterall 1971) and his classification system is described in detail in chapter 7.2.1.2. The lateral third of the femoral head was early recognized to have predictive value for the final shape of the femoral head (Kemp and Boldero 1966) which provided the basis for the development of the lateral pillar classification system (Herring et al. 1992, Herring et al. 2004) (for detailed description see chapter 7.2.1.2). Numerous authors have shown that Herrings original 3-group classification assessed at fragmentation is a strong predictor for the final radiographic outcome (Herring et al. 1992, Ritterbusch et al. 1993, Farsetti et al. 1995, Ismail and Macnicol 1998). The Herring group later reassessed the predictive value of their own classification and introduced an intermediate group (Herring et al. 2004), changing their 3- group classification into a 4-group classification. The authors stated that the modified lateral pillar classification was a reliable predictor of the final radiographic outcome, assessed by the Stulberg classification. Our intention with the present studies was to assess whether the modification of the lateral pillar system would be as reliable and predictive compared to the original classification and the Catterall classification.

4.5.3.2 Femoral head subluxation

Subluxation of the femoral head has been associated with a poor radiographic outcome. Several parameters have been proposed to describe the gradual decrease of the femoral head coverage, including head-socket distance (Waldenström 1922), Acetabular-head quotient (Heyman and Herndon 1950), Centre-edge angle (Wiberg and Frey 1939) and a
combination of these measurements (Danielsson et al. 1982). It has been stated that gradually increased uncoverage of the lateral part of the femoral head subsequently increases the risk of femoral head deformity, since the molding acetabular template with essential counter pressure is missing in the proliferating reossification phase.

4.5.3.3 Acetabular dysplasia

The prognostic value of acetabular changes has been investigated sporadically. Hypertrophy and bicompartamentalisation was reported to cause extrusion of the femoral head with subsequent subluxation and inferior radiographic outcome (Yngve and Roberts 1985, Joseph 1989). It also has been noted that acetabular dysplasia with increased obliquity and reduced concavity led to a poor outcome (Bellemans et al. 1979). Grzegorzewski et al. (2006) showed that a deformed lateral acetabular margin was associated with a poor radiographic outcome and the hypothesized that this was due to reduced lateral support of the femoral head.

Compared to the abundance of studies assessing the predictive value of femoral head changes, acetabular changes have been of little interest within Perthes research. There are to date no prospective studies investigating the role of the acetabulum on the final radiographic outcome. We were interested in whether acetabular changes might influence the femoral head shape and which relation they have compared to other prognostic femoral head changes.

4.5.4 Long-term prognosis

The first long-term follow-up reports indicated a benign course of the disease. Gower and Johnston (1971) reviewed 36 patients with an average age of 45 years. They found that 86% of the hips were still functioning and only 3 had been treated with arthroplasty. Catterall (1982) stated that 58% of his patients got better without any treatment, whereas the
remaining third improved with time. McAndrew and Weinstein (1984) followed up on the 36 patients presented initially by Gower and Johnston. 40% of the Perthes hips had undergone arthroplasty and 20% had developed a decreased function, which indicated a marked decline of hip survival between the 4th and 5th decades of life. Later, a number of studies confirmed deterioration of hip function and development of osteoarthritis in the 5th decade (Lecuire 2002, Heesakkers et al. 2015) especially when compared to a match pair control group (Froberg et al. 2011). In a landmark study, Stulberg et al. (1981) investigated the relationship between the residual hip deformity at maturity and the further development of the hip. They reported that femoral head flattening was highly correlated with secondary osteoarthritis in adulthood. The femoral head sphericity and the hip joint congruency at time of healing has since then been widely accepted as determining factors for the long-term hip survival (Mose 1980, Yrjonen 1999, Heesakkers et al. 2015).
5 Aims of study

The aims of the present study were to

1. Describe acetabular changes that occur during the course of Perthes disease

2. Evaluate the interobserver agreement and reliability of radiographic acetabular measurements in Perthes disease

3. Investigate the predictive value of acetabular changes on radiographic outcome 5 years after diagnosis

4. Describe the natural course and determine specific prognostic factors of bilateral Perthes disease

5. Evaluate the interobserver agreement and prognostic value of radiographic classifications assessing the extent of femoral head necrosis at an early stage of the disease
6 Summary of papers I - IV

6.1 Paper I

The Acetabulum in Perthes’ disease: Inter-observer agreement and reliability of radiographic measurements

**Background and purpose** – Perthes’ disease leads to radiographical changes in both the femoral head and the acetabulum. We investigated the inter-observer agreement and reliability of 4 radiographical measurements assessing the acetabular changes.

**Patients and methods** - We included 123 children with unilateral involvement, femoral head necrosis more than 50% and age at diagnosis 6 years or older. Radiographs were taken at onset, 1 year and 5 years after diagnosis. Sharp’s angle, Acetabular depth-width ratio (ADR), lateral acetabular inclination (LAI) and acetabular retroversion (Ischial-Spine-Sign, ISS) were measured by 3 observers. Before measuring, 2 of the observers had a consensus meeting.

**Results** - We found good agreement and moderate to excellent reliability for Sharp’s angle for all observers (ICC> 0.80 with consensus, ICC 0.46 – 0.57 without consensus). There was good agreement and substantial reliability between observers with consensus meeting for ADR (ICC 0.62 – 0.89). Low levels of agreement and poor reliability were found for observers without consensus meeting. LAI showed fair agreement throughout the course of the disease (kappa 0.28- 0.52). The agreement between observations for ISS was ranging from fair to good (kappa 0.20 – 0.76).

**Interpretation** - Sharp’s angle showed highest reliability and agreement throughout the course of the disease. ADR was only reliable and obtained good agreement between the
observers when landmarks were clarified prior to measuring the radiographs. Thus, we recommend both parameters in clinical practice, provided a consensus is established for ADR. The observations for LAI had only fair agreement and ISS showed inconclusive agreement in our study. Thus, LAI and ISS can hardly be recommended in clinical practice.

6.2 Paper II

The acetabulum in Perthes’ disease. A prospective study of 123 children

Background and purpose - We assessed the radiographic changes of the acetabulum during the course of Perthes’ disease and investigated whether they were associated with femoral head sphericity 5 years after diagnosis.

Patients and methods - We studied 123 children with unilateral Perthes’ disease, femoral head necrosis more than 50 % and age at diagnosis 6 years or older. Pelvic radiographs were taken at onset, 1 year and 5 years after diagnosis. Sharp’s angle, acetabular depth-to-width ratio (ADR) and lateral acetabular inclination were measured.

Results - Compared to the unaffected hips, the Perthes’ hips developed significantly higher Sharp’s angles (p \(0.001\)) and a higher proportion with an upward-sloping lateral acetabular margin (Perthes’ hips: 49 %, unaffected hips 1 %). The mean ADR values were significantly lower on the affected side at all stages (p \(0.001\)). ADR values at diagnosis were associated with a more spherical femoral head at the 5-year follow-up (odds ratio (OR) 1.012, 95 % confidence interval (CI) 1.002–1.022, p = 0.016). None of the other acetabular parameters were significantly associated with the femoral head shape 5 years after diagnosis.
**Interpretation** - The acetabulum developed an increasingly dysplastic shape in the course of Perthes’ disease. Early dysplastic changes of the acetabulum were not associated with a poor radiological outcome 5 years after diagnosis. Routine measurement and monitoring of acetabular changes in plain radiographs were of little prognostic value and can, therefore, hardly be recommended in clinical practice.

6.3 Paper III

**Outcome and prognostic factors in bilateral Perthes disease**

**A PROSPECTIVE STUDY OF 40 CHILDREN WITH 5 YEARS FOLLOW-UP**

**Background and purpose** - The aims of this prospective study were to describe the course of non-operatively treated, bilateral Perthes disease, and to determine specific prognostic factors for the radiographic outcome.

**Patients and methods** – We identified 40 children with a mean age 5.9 years (1.8 to 13.5), who were managed non-operatively for bilateral Perthes disease from our prospective, multicentre study of this condition, which included all children in Norway who were diagnosed with Perthes disease in the 5-year period between 1996 and 2000. All children were followed up for 5 years. The hips were classified according to the Catterall classification. A modified 3-group Stulberg classification was used as outcome measure, with spherical femoral head as good outcome, oval head as fair, and flat femoral head as poor outcome.

**Results** – Concurrent, simultaneous bilateral Perthes disease was seen in 23 children and 17 had a sequential onset of bilateral disease. The mean delay in onset for the second hip in
the latter group was 1.9 years (0.3 to 5.5). The 5-year radiographic outcome was good in 30 (39\%), fair in 25 (33\%), and poor in 21 (28\%) of the hips. The strongest predictors of poor outcome were > 50% femoral head, with odds ratio (OR) 19.6, and age at diagnosis > 6 years (OR 3.3). Other risk factors for poor outcome were timing of disease onset, where children with sequential onset had higher risk than those with concurrent onset of bilateral disease (p=0.021, chi-squared test). Following a diagnosis of Perthes disease in one hip, there was a 5\% chance of developing it in the contralateral hip.

**Interpretation** - Our results imply that we need to distinguish between children with concurrent onset and those with sequential onset of bilateral Perthes disease, as the outcome may be different. This has not been previously described. Children with concurrent onset of bilateral disease had similar outcome to our previous series of those with unilateral disease, whereas children with sequential onset disease had worse prognosis. The increased risk of developing Perthes disease in the contralateral hip in those with unilateral disease is important information to the child and parents.

6.4 Paper IV

**Radiographic classifications in Perthes disease**

**Interobserver agreement and association with femoral head sphericity at 5-year follow up**

**Background and purpose** - Different radiographic classifications have been proposed for prediction of outcome in Perthes disease. We assessed whether the modified lateral pillar classification would provide more reliable interobserver agreement and prognostic value compared with the original lateral pillar classification and the Catterall classification.
Patients and methods - 42 patients (38 boys) with Perthes dis- ease were included in the interobserver study. Their mean age at diagnosis was 6.5 (3–11) years. 5 observers classified the radio- graphs in 2 separate sessions according to the Catterall classification, the original and the modified lateral pillar classifications. Interobserver agreement was analysed using weighted kappa statistics. We assessed the associations between the classifications and femoral head sphericity at 5-year follow-up in 37 non-operatively treated patients in a crosstable analysis (Gamma statistics for ordinal variables, γ).

Results - The original lateral pillar and Catterall classifications showed moderate interobserver agreement (kappa 0.49 and 0.43, respectively) while the modified lateral pillar classification had fair agreement (kappa 0.40). The original lateral pillar classification was strongly associated with the 5-year radiographic outcome, with a mean γ correlation coefficient of 0.75 (95% CI: 0.61–0.95) among the 5 observers. The modified lateral pillar and Catterall classifications showed moderate associations (mean γ correlation coefficient 0.55 (95% CI: 0.38–0.66) and 0.64 (95% CI: 0.57–0.72), respectively).

Interpretation - The Catterall classification and the original lateral pillar classification had sufficient interobserver agreement and association to late radiographic outcome to be suitable for clinical use. Adding the borderline B/C group did not increase the interobserver agreement or prognostic value of the original lateral pillar classification.
7 General Discussion

7.1 Patients

7.1.1 Nationwide study on Perthes disease in Norway (Paper I-III)

The Norwegian Pediatric Orthopedic Society initiated a nationwide study on Perthes disease in January 1996. Over a 5-year period all hospitals with pediatric orthopedic service in Norway reported all new incidences of Perthes disease. The local orthopedic surgeon established the diagnosis based on clinical examination and characteristic radiographic findings on AP pelvis and frog-leg lateral radiographs. A total of 425 children were included in this prospective study. Radiographs were taken at diagnosis, and at 1 and 5 years after diagnosis. Based on AP and frog-leg lateral view, the affected hips were classified according to the Catterall classification (Catterall 1971) and lateral pillar classification (Herring et al. 1992).

For Paper I and II, one of the objectives was to examine the acetabular development of affected hips and compare the observed changes with the non-affected, normal side. Hence, it was necessary to exclude all bilateral cases (n=55). In younger children, the acetabular walls as well as the triradiate cartilage are not ossified. These structures cannot be identified on standard radiographs, which limits the interpretation of acetabular development in this cohort. Thus, we chose to include all children with an age at onset of 6 years and older (n=323). Acetabular changes may be subtle in Perthes hips with mild affection. We therefore included only hips with a higher degree of initial hip involvement (more than 50 % femoral head necrosis) since more pronounced changes could be expected (n=152). 29 patients had to be excluded due to inadequate exposure of acetabular landmarks. In total, 123 children were eligible for analysis in Paper I and II.
For Paper III, we identified a subgroup of 55 children with bilateral involvement. We excluded 5 patients who had been diagnosed with Perthes disease in one of their hips prior to the study period. The objective of this paper was to study the natural course of bilateral affected hips in Perthes disease. There is strong evidence in the literature that operative treatment may alter the course of Perthes and therefore we excluded 8 patients due to operative treatment (7 patients with unilateral femoral osteotomies and 1 patient with bilateral femoral osteotomies). 2 patients with inadequate radiographs were not eligible for radiographic analysis and were excluded. Thus, a total of 40 patients were included in Paper III.

7.1.2 Retrospective review of hospital radiographic archive (Paper IV)

By a systematic search of the radiographic archive of our institution, we identified 152 children who had been treated for Perthes disease between 1950 and 1984. 139 children had adequate radiographic follow-up with good radiographic quality at least 5 years after diagnosis. We selected a random sample of 50 patients using a random-number regenerator. One objective of paper IV was to classify the hips according to Herring’s original lateral pillar classification (Herring et al. 1992). This can only be done reliably when the contralateral side is unaffected; hence 5 patients with bilateral involvement were excluded. We excluded further 3 patients due to advanced Waldenström staging (reossification stage) at initial consultation. Thus, 42 patients were included in Paper IV.
7.2 Methods

7.2.1 Radiographic imaging

At time of inclusion of our patients (all papers), conventional radiography was the basis of all relevant classification systems assessing the extent of femoral head necrosis as well as the late radiographic outcome in Perthes disease. MRI and CT were not available at all institutions at that time and advanced diagnostic imaging had been reserved for special indications. All 4 included papers are therefore based on conventional radiography and AP pelvis and frog-leg lateral radiographs were assessed at time of diagnosis, 1 and 5 years after diagnosis. Radiographs from the hospital archive (Paper IV) as well as early radiographs in the nationwide study (until 2001, Paper I-III) were mainly projected onto film. Imaging shortcomings such as low brightness, lack of contrast and small scale developing led to substandard visual quality in some cases. Hence, a number of radiographs had to be excluded in all 4 articles.

Paper I – III is based on a nationwide study where a large number of institutions have been involved. We have not been able to standardize the imaging process of the participating radiographic departments concerning imaging quality and patient positioning. Patient positioning is crucial in order to obtain AP pelvic radiographs with standardized tilt and symmetric rotation of the pelvis. Siebenrock et al.(2003) outlined criteria for symmetric rotation and tilt such as symmetric appearance of the obturator foramina and the tip of the coccyx pointing toward the symphysis pubis in a defined distance. Asymmetrical projection of the obturator foramina with pelvic rotation along the longitudinal axis toward the affected hip may be caused by ipsilateral muscular atrophy (Schiller and Axer 1972). It is known that Perthes disease may lead to muscular atrophy over time and we observed an increased number of radiographs with malrotation in the later stages of the disease (1 year
and 5 year after diagnosis) (Paper I). Other authors focusing on standardized criteria of AP pelvic radiographs in Perthes disease had similar dropout rates as observed in our study (Larson et al. 2011, Kawahara et al. 2012). Hence, unstandardized patient positioning as well as increasing muscular atrophy in the course of Perthes disease may have contributed to asymmetrical pelvic rotation. Another contributing factor leading to exclusion of radiographs was the shielding of essential landmarks especially on the acetabular side (Paper I and II).

### 7.2.1.1 Acetabular measurements

In paper I and II we assessed the development of the acetabulum based on radiographic changes in the course of the disease. We selected parameters that measured the acetabular version, inclination and shape.

*Acetabular inclination*

Many parameters have been described to assess the acetabular inclination on AP pelvic radiographs. We selected Sharp’s angle (Sharp 1961) since this parameter can be measured both in skeletally mature and immature patients. This angle was first described in the assessment of hip dysplasia. A reference line was drawn between the inferior points of the teardrops on AP pelvic radiographs. The angle was formed by this reference line and a line connecting the inferior point of the teardrop and lateral edge of the acetabulum. The teardrop is a non-anatomical structure, which results of an overprojection of the fossa acetabuli on the lateral side and the ilio-ischial line on the medial side. The characteristic shape of the teardrop is dependent on a centered x-ray beam on AP pelvic radiographs. Substandard imaging technique may therefore influence the projection of the teardrop, making it difficult to delineate the lower border and hence jeopardizing the landmarks of both the reference line as well as the articular line. Conflicting results concerning the
influence of pelvic rotation and tilt on Sharp’s angle have been published. While Henebry et al. (2013) and Lee et al. (2011) found that Sharp’s angle did not respond to changes of pelvic positioning, Monazzam et al. (2013) observed a significant change of Sharp’s angle with increasing pelvic rotation. Since the latter study was published after commencing our Paper I and II, we did not define symmetric pelvic rotation and tilt as a basic requirement to measure the angle.

![Image of pelvic anatomy with annotations](image)

**Figure 6.** Sharp’s angle (left hip), the acetabular depth (D) to width (W) ratio (ADR, right hip) and the Ischial Spine Sign (ISS) are depicted on a schematic illustration of an AP pelvis radiograph

**Acetabular shape**

The acetabulum is a cup-shaped anatomical structure and it poses obvious limitations to describe its shape on a 2D projection on AP pelvic radiographs. However, the relation between the acetabular length and width has been studied as an indirect marker of acetabular shape especially in the assessment of hip dysplasia. Reference values of the acetabular depth to width ratio (ADR) in anatomically normal hips (Bellemans et al. 1979)
and in hip dysplasia have been established to define normal acetabular development (Novais et al. 2016) versus a dysplastic alteration in several age groups (Cooperman et al. 1983). The acetabular depth and width were measured on AP pelvic radiographs as described by Heyman and Herndon (1950) (Figure 6). The width was defined as the length of a line connecting the upper osseous acetabular margin and the lower end of the teardrop. This landmark is often more accurately defined than the lower acetabular margin, especially in younger patients (Heyman and Herndon 1950). The depth was defined as the distance from the width line to the deepest point of the acetabulum. We defined the ADR according to Cooperman et al.(1983) as depth/width × 1,000. Since one of the landmarks is the lower margin of the teardrop, the ADR is as vulnerable to substandard imaging technique as the Sharp’s angle. To our knowledge, there is no literature examining the relationship between this parameter and the pelvic rotation and tilt.

Retroversion

Assessment of the anatomy of a 3-dimensional object such as the acetabulum on a 2-dimensional radiograph has obvious limitations. However, the crossover sign (Jamali et al. 2007) and the Ischial Spine Sign (ISS) (Kalberer et al. 2008) (Figure 6) on AP pelvic radiographs have been found to diagnose acetabular retroversion reliably. The acetabular walls are not ossified in younger patients, which limits the use of the crossover sign in the pediatric patient population. The ischial spine is an anatomical structure that extends from the posterior border of the ischium, somewhat inferiorly to the acetabulum. It can be detected reliable also in children as well. Due to bony overprojection, it is barely visible on AP pelvic radiographs with symmetric rotation and standardized tilt, given that the acetabulum has a normal version. With increasing retroversion, the ischial spine becomes more prominent as a triangular structure medial to the pelvic rim. The ISS is sensitive to
patient positioning. Pelvic rotation along the longitudinal axis toward the examined hip and/or inlet-tilt of the pelvis lead to a prominence of the ischial spine without existing acetabular retroversion (Kakaty et al. 2010). A substantial number of AP pelvic radiographs had therefore to be excluded in Paper I, since they did not meet the criteria of symmetric rotation and tilt.

*Lateral acetabular inclination*

The bony lateral acetabular margin and its labrum are vital stabilizing hip structures, supporting the femoral head from the weight-bearing dome further laterally. The importance of the lateral acetabular margin has been recognized especially in hip dysplasia as many indices such as the angle of Wiberg (1939), the extrusion index and the acetabular index use this landmark in the hip assessment. The inclination of the lateral acetabular margin and its role for hip stability in hip dysplasia was first introduced by Cooperman et al. (1983) and later applied to children with Perthes disease by Grzegorzewski et al. (2006). The authors categorized the slope of the lateral margin in 3 groups, depending on whether the lateral lip of the acetabulum was below, horizontal or above the weight-bearing dome of the acetabulum.

![Diagram](image)

Figure 7. Classification of the slope of the lateral acetabular margin. Group a: the lateral lip of the acetabulum (LLA) is below the weightbearing dome (dotted line); Group b: LLA is at
the same level as the weightbearing dome; Group c: LLA is above the weightbearing dome of the acetabulum.

In younger children, most of the upper lateral acetabular margin is not ossified and hence cannot be detected correctly on standard radiographs. We considered this in our study design and excluded all patients younger than 6 years from the analysis in Paper I and II.

7.2.1.2 Radiographic Classifications

Waldenström staging

In all papers, the radiographic phases were determined for the affected hips at onset, 1-year and 5-year follow-up according to Waldenström (1922). The initial phase is characterized by differences in epiphyseal height, width and bony density compared to the normal hip. The fragmentation phase included hips where the epiphysis was fragmented and necrotic bone was partly or totally resorbed. Hips with signs of new bone formation were classified as being in the reossification phase, and those that had been fully rebuilt and recovered were classified as being in the healing stage.

Catterall classification

In 1971, Catterall introduced a 4-group classification (Catterall 1971) based on the extent of femoral head involvement on AP pelvis and frog-leg lateral projections. In paper I-IV, radiographs at the time of diagnosis and 1-year follow-up were reviewed and only those that showed greatest involvement of the femoral head in the fragmentation phase were selected. In group 1, only the anterior part (25%) of the epiphysis is involved. The epiphysis is
not collapsed and complete absorption of the involved segment is without sequestrum formation.


Hips with involvement up to 50 % of the epiphysis, and formation of a dense collapsed segment or sequestrum are classified in group 2. In the AP pelvic radiographs there are viable lateral and medial fragments, maintaining the epiphyseal height in case of central collapse.

Group 3 included hips where more than 50% and less than 75% of the femoral head was affected on frog-leg lateral projections. Only the most posterior part remains uninvolved.


In group 4, the whole epiphysis is affected, with total collapse on AP pelvic radiographs and bony necrotisation of all parts on frog-leg lateral projection.


Catterall described additionally “head at risk signs” (Catterall 1971). These are the Gage’s sign (small osteoporotic segment which forms a transradiant “V” on the lateral side of the
epiphysis), calcification lateral to the epiphysis, lateral subluxation with increase of medial joint space, and a transverse growth plate seen on AP view. Since the literature differs substantially on the prognostic value regarding the “head at risk” signs (McAndrew and Weinstein 1984, Mukherjee and Fabry 1990, Fulford et al. 1993), we did not evaluate these in our studies.

In papers I-III, we modified the Catterall 4-group classification, combining Catterall groups 1 and 2 (less than 50% femoral head necrosis), and Catterall groups 3 and 4 (more than 50% of femoral head necrosis). Based on previous studies on the same patient material, this simplification of the Catterall classification was more reliable (Wiig et al. 2002) and a stronger predictor of late radiographic outcome (Wiig et al. 2008) compared with the original 4-group classification.

*Original lateral pillar classification*

Herring et al.(1992) introduced a 3-group classification based on the height of the lateral portion of the femoral epiphysis (termed lateral pillar) on AP pelvis radiographs. It has to be applied in fragmentation phase and it compared the height of the lateral pillar with the unaffected side on AP pelvic radiographs. Group A hips had no loss of height of the lateral pillar. Group B hips had increased lucency and loss of height, but not exceeding 50%. Group C hips exhibited more lucency and > 50% loss of height. In papers I, II and IV, we applied this classification when the hips were in fragmentations phase with a maximum degree of involvement.
Figure 12. The original lateral pillar classification. In group A the lateral pillar has normal height maintained; in group B the height is depressed but > 50% is maintained; in group C < 50% of height maintained. Reprinted from “The lateral pillar classification of Legg-Calve-Perthes disease.” by J. A. Herring, 1992, J Pediatr Orthop; 12 (2): 143-50. Copyright 2008 by Wolters Kluwer Health, Inc. Reprinted with permission.

*Modified lateral pillar classification*

Later, the Herring group reassessed their original study (Herring et al. 2004). They identified a group of hips with more severe radiographic findings than those seen in group B but less severe than those typical in group C. Thus, they introduced a new group termed B/C borderline, transforming their original 3-group classification into a classification with 4 categories. We did not apply the B/C borderline group in papers II-III as this group was introduced by Herring et al. several years after the inclusion and radiographic analysis of this study.
Figure 13. The borderline B/C group of the modified lateral pillar classification. Herring et al. defined the “radiographic findings in this group as (I) a very narrow lateral pillar (2 to 3 mm wide) that was > 50% of height, (II) a lateral pillar with very little ossification but at least 50% of the original height, or (III) a lateral pillar with exactly 50% of original height that is depressed relative to the central pillar” (Herring et al. 2004).

Stulberg classification

Stulberg et al. (1981) proposed a 5-group classification to describe the residual hip joint deformity and the onset of degenerative joint disease after healed Perthes disease. Neyt et al. (1999) refined Stulberg’s original classification and we used this algorithm since it defines more explicit criteria to categorize the hips. Spherical hips with normal femoral head, neck and acetabulum were classified in Class I. Class II hips had a spherical femoral head with either coxa magna, short neck, or steep acetabulum. If the femoral head was round, but the contour did not fit within a 2 mm concentric circle template on both the AP and frog-leg lateral view, it was defined as ovoid and categorized as class III. In class IV hips, at least 1/3 of the femoral head contour resembled to a straight line in one of the projections and they
were defined as flat. However, the acetabulum and the femoral head remained congruent in this category. Class V hips had flat femoral head with an incongruent hip joint.

The reliability of Stulberg’s 5-group classification is debated in literature (Neyt et al. 1999, Agus et al. 2004). Based on previous results of our study group (Wiig et al. 2007), the reliability of the Stulberg classification was improved by modification into a 3-group classification, without jeopardizing its prognostic value. The grouping was based on the shape of the femoral head alone, where all spherical heads were merged into Group A, ovoid femoral heads were classified as Group B and flat femoral heads were classified as Group C. We used the modified Stulberg 3-group classification in paper II-IV as a measure of final radiographic outcome in healed Perthes hips.

7.2.2 Statistical analysis

7.2.2.1 Interobserver analysis

Continuous data

Several statistical strategies have been described in the evaluation of reliability and agreement in measurement studies for numerical data (Bland and Altman 1986, 1999, Petrie 2006). Reliability parameters (e.g. intraclass correlation coefficient) are related to how well different measurements can be distinguished from each other, while agreement parameters like the Bland Altman approach assess how close the scores for repeated measurements are. To answer the question whether a parameter is reproducible, both reliability and agreement have to be tested. Therefore we chose to include both the Bland Altman method and intraclass correlation coefficient (ICC) in paper I. The ICC was analyzed using a one-way random effect model assuming a single measurement (McGraw and Wong 1996). ICC of 0 implies no more reliability than would be expected by chance alone, whereas values close to
1 indicates perfect reliability. We interpreted the intermediate values according to the guidelines described by Landis and Koch (1977): values < 0.01 indicate poor reliability; 0.01 - 0.20 slight reliability; 0.21 - 0.40, fair reliability; 0.41 - 0.60, moderate reliability; 0.61 - 0.80, substantial reliability; and > 0.81, excellent reliability.

The Bland Altman approach was used to examine the interobserver agreement (Bland and Altman 1986, 1999). It assesses the differences of numerical data between a pair of observers in relation to their mean values. We calculated the differences between the radiographic measurements obtained by 2 observers for each radiograph and calculated the mean differences and their respective standard deviation. Good agreement is defined when mean differences are below 5 % of their respective mean values. The 95% limits of agreement were estimated as the mean difference between the 2 measurements +- 1,96 standard deviation (SD). This range includes 95% of the inter-observer differences.

**Categorical data**

In paper I and IV we assessed the interobserver agreement by calculating the weighted kappa (Cohen 1968) for each pair of observers. Kappa statistics with linear weighting was used, defining the relative distance between ordinal categories as 1 (Lowry 2015). The average of those kappa values for each classification was recorded as the overall kappa value (Light 1971). Kappa values range from -1 to 1, with 1 indicating perfect agreement, 0 indicating random agreement, and -1 indicating complete disagreement. As suggested by Landis and Koch (1977), we interpreted the intermediate kappa values as follows: values of < 0,20 indicate poor agreement; 0,21 - 0,40 fair agreement; 0,41 - 0,60 moderate agreement; 0,61 - 0,80 good agreement and values > 0,80 excellent agreement.

We demonstrated the statistical complexity of interobserver studies in paper IV where we reassessed a published study (de Billy et al. 2002), examining the interobserver agreement of
the Catterall 4-groups classification. The authors presented excellent interobserver agreement using ICC statistics (ICC=0.94; (de Billy et al. 2002)), without stating sufficiently the variant of ICC which was used (missing unit and effect of ICC). We reanalyzed the ICC with the given raw data in the article, based on single ratings proved by a single observer, which decreased the ICC values from excellent to good. Further the raw data were analyzed using linear weighted kappa statistics, yielding 36 pairs of observations and an average kappa value of 0.54 (moderate agreement, according to Landis and Koch). The reassessment illustrated to which degree the result differed, depended on the chosen statistical approach. This was in accordance with Norman and Streiner (2008), who showed that using ICC (two-way, mixed, single-measures, consistency) was identical to a quadratic weighted kappa which resulted in higher kappa values as with normal linear weighting and hence implied a higher degree of interobserver agreement.

7.2.2.2 Predictor of final radiographic outcome

Continuous data

In paper II and III, we evaluated the numerical parameters at time of diagnosis regarding their association with the radiographic outcome at 5-year follow-up. All parameters were run in simple models, only estimating the effect of the single parameter one at a time. Outcome was the modified 3-group Stulberg classification. The continuous data were analyzed using a one-way analysis of variance (ANOVA).

Categorical data

In paper IV, we assessed the value of the Catterall and Herring classification to predict the final radiographic outcome. Several observers classified the healed Perthes hips and we defined the Stulberg classification for each patient, which had been assigned by the majority
of the observers. The intention of this algorithm was to eliminate all uncertainties regarding 
the reliability of the final radiographic outcome, and we termed the outcome accordingly the 
“true” Stulberg 3-group classification. The prognostic value of the Catterall, the modified and 
the original lateral pillar classification was assessed in a cross table analysis with “true” 
Stulberg 3-group classification as outcome variable (n=37). Gamma statistics for ordinal 
variables were used and the values interpreted as follows: values of < 0.24 no association, 
0.25 - 0.49 weak; 0.50 - 0.74 moderate and values > 0.74 strong association.

*Multivariable regression analysis*

In paper II and III, the multivariable regression analysis were performed in collaboration with 
a statistician. In paper II, an ordinal regression model was used to estimate the cumulative 
ods ratio (OR) for worse radiological outcome (modified Stulberg classification) with 
selected predictor variables. The following variables were each analyzed separately in the 
ordinal regression model: age at onset of the disease, gender, Catterall classification, lateral 
pillar classification, received treatment, femoral head coverage at 1-year follow-up and ADR. 
Further, those that were significantly associated with the Stulberg 3-group classification 
(with p< 0.05 as criteria) were included in the final multivariable ordinal regression model to 
assess their predictive ability regarding worse radiological outcome. Results were regarded 
as statistically significant if the p- values were below 0.05.

In paper III we assessed patients with both hips affected. In order to account for 
dependencies within each patient, ordinal regression was performed using Generalized 
Estimating Equations (GEE) with an exchangeable working correlation matrix in paper III. The 
results from the ordinal regression were presented as cumulative odds ratios with 95% 
confidence intervals and p-values.
7.3 Results

7.3.1 Acetabular changes during the course of Perthes disease

In paper II, we assessed the radiographic changes of the acetabulum from diagnosis to 5 years after diagnosis. The acetabulum developed a more dysplastic shape, with steeper inclination, loss of concavity and a lateral upward sloping margin. These changes were seen at different points of time. The acetabular inclination was almost similar at the time of diagnosis in both the affected and unaffected side. Later the affected hips failed to incline compared to their contralateral side, with stationary Sharp’s angles throughout the course. There was a higher proportion of hips with an upward sloping lateral margin, and this observation was most pronounced at the 5-year follow-up. To our knowledge, only 1 previous study has investigated the acetabular inclination in Perthes disease. Joseph (1989) found that there was a marked difference in Sharp’s angle in affected hips compared to the unaffected side in skeletally immature patients. The authors did not specify at which stage of the disease they measured the inclination, which made direct comparison with our study difficult. However, the author presented higher Sharp’s angle in symptomatic adult patients with Perthes disease (40.1° in affected hips and 36.1° in unaffected hips), which is in accordance with our results at 5-year follow-up. Grzegorzewski et al. (2006) observed a higher proportion of upward sloping lateral acetabular margins at fragmentation, which is similar to our results at diagnosis. The progression of a steep slope until 5 years after diagnosis as observed in our patients was not as marked. In contrast to Sharp’s angle and the lateral acetabular inclination, we noticed that changes of the acetabular depth and width were most pronounced within the first year after diagnosis and thereafter the acetabula remodeled gradually to a more normalized shape. It is still unclear whether acetabular changes occur primarily, linked to the pathology itself, or secondary to an enlarged femoral
head. Although the concept of mechanical induced acetabular adaptation has been recapitulated in most of the literature, there is some evidence to support non-mechanical linked alteration. Increased metabolic activity in the area of the triradiate cartilage has been reported by several authors (Joseph 1989, Madan et al. 2003). It was speculated that hyperemia might lead to increased growth at the triradiate cartilage, causing a widening of the acetabulum. Disproportional growth due to hyperemia in close relation to the growth plate is a phenomenon also known in other conditions such as juvenile idiopathic arthritis (Breton et al. 2012). Radiographic appearance of a wider acetabulum early in the course of Perthes disease as seen in paper II, had been observed by other authors as well (Bellemans et al. 1979). This could imply that enlarged acetabular dimensions are not exclusively caused by the mechanically induced forces from an enlarged femoral head. Several authors found a thickening of medial cartilage of the acetabulum due to inflammatory processes and increased metabolism (Gershuni et al. 1981, Yngve and Roberts 1985), leading to hypertrophy and reduced acetabular depth, independent of mechanical forces. Joseph (1989) suggested that increased appositional growth was mediated by synovitis and hyperemia, being a factor contributing to depth reduction.

7.3.2 Interobserver agreement and reliability

7.3.2.1 Sharp’s angle, ADR, slope of lateral margin and ISS

We assessed the interobserver agreement and reliability of 4 commonly used acetabular measurements in paper I. Our results showed good interobserver agreement and acceptable reliability for Sharp’s angle and ADR, while the lateral acetabular inclination and acetabular retroversion (ISS) were not suitable in Perthes disease due to low or diverging interobserver agreement.
To our knowledge, Sharp’s angle and ADR have not been validated previously in Perthes disease. In DDH, Sharp’s angle have been found to be a reliable measurement with a high degree of interobserver agreement in skeletally mature (Nelitz et al. 1999, Engesaeter et al. 2012) and immature patients (Agus et al. 2002). We found good interobserver agreement for Sharp’s angle at all stages of Perthes disease. The reliability was substantial to excellent in the earlier phase, and moderate at 5 years after diagnosis. Our results could confirm that Sharp’s angle is a reliable measurement with high interobserver agreement in Perthes disease.

Interobserver reliability and agreement of ADR has been assessed in children with DDH, but the findings differed widely. Takatori et al. (2010) reported large values for the coefficient of variation for ADR, indicating a low degree of agreement, which was in accordance with the results of Clohisy et al.(2009) who demonstrated fair reliability for acetabular depth in adults. In contrast, Nelitz et al. (1999) found moderate interobserver reliability for ADR with ICC values ranging from 0.58 to 0.63. Engesaeter et al.(2012) showed substantial interobserver reliability for ADR (ICC = 0.77) in patients with DDH. Both publications reported good interobserver agreement. Our results showed divergent interobserver agreement and reliability depending on the stage of Perthes disease. In the early stages the interobserver agreement was good and substantial reliability could be found. At 5-year follow up there were lower but still expectable levels of interobserver agreement and reliability.

The reliability and agreement of the lateral acetabular inclination and the ISS has not been previously validated in Perthes disease. However, there is one study assessing the interobserver agreement of the ISS in another condition. Kappe et al. (2011) emphasized that the ISS was subject to considerable interobserver differences in skeletally mature
patients, which in accordance with our results with interobserver agreements ranging from fair to moderate.

7.3.2.2 Catterall and lateral pillar classification

In paper IV, we assessed the interobserver agreement of the Catterall classification, the original lateral pillar classification and Herring’s modification of the lateral pillar classification. We found that the Catterall and the original 3-group lateral pillar classification had moderate interobserver agreement, while the modified 4-group lateral pillar classification had fair interobserver agreement.

Several authors have investigated the interobserver agreement for the Catterall and original lateral pillar classification. The studies differ substantially in the number of observers, assessed radiographs and statistical approach, hence direct comparison of these results with paper IV should be done with caution. An overview of the results is given in tables 2 and 3, including the results of paper IV.
Table 2. Interobserver agreement of the Catterall 4-group classification in 6 previous studies

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Obs</th>
<th>Statistics</th>
<th>Mean (Range)</th>
<th>Interpretation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pietrzak et al. (2004)</td>
<td>63</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.39 (0.28 – 0.42)</td>
<td>Fair agreement</td>
</tr>
<tr>
<td><strong>Present study</strong></td>
<td>42</td>
<td>5</td>
<td><strong>Weighted kappa</strong></td>
<td><strong>0.43 (0 – 0.73)</strong></td>
<td><strong>Moderate agreement</strong></td>
</tr>
<tr>
<td>Sambandam et al. (2006)</td>
<td>44</td>
<td>2</td>
<td>Weighted kappa</td>
<td>0.44</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>de Billy et al. (2002)</td>
<td>19</td>
<td>9</td>
<td>Weighted kappa +</td>
<td>0.54 (0.36 – 0.77)</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>Wiig et al. (2002)</td>
<td>63-158</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.49 – 0.62</td>
<td>Moderate to good agreement</td>
</tr>
<tr>
<td>Simmons et al. (1990)</td>
<td>40</td>
<td>15</td>
<td>Weighted kappa</td>
<td>0.55 (0.49 – 0.64)</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>Christensen et al. (1986)</td>
<td>100</td>
<td>4</td>
<td>Weighted kappa</td>
<td>0.62 (0.50 – 0.67)</td>
<td>Good agreement</td>
</tr>
</tbody>
</table>

n is number of patients; Obs is number of observers, * Interpretation of kappa values (Landis and Koch 1977): < 0.2 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 good agreement; + recalculated with data given in the original article
Table 3. Interobserver agreement of Herrings original lateral pillar (3-group) classification in 6 previous studies

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Obs</th>
<th>Statistics</th>
<th>Mean (Range)</th>
<th>Interpretation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study</td>
<td>42</td>
<td>5</td>
<td>Weighted kappa</td>
<td>0.49 (0.35 – 0.72)</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>Podeszwa et al.</td>
<td>33</td>
<td>5</td>
<td>Cohen’s kappa</td>
<td>0.51 (0.43 – 0.62)</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>(2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herring et al.</td>
<td>32</td>
<td>16</td>
<td>Kappa, unspecified</td>
<td>0.52</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>(1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akgun et al.</td>
<td>50</td>
<td>3</td>
<td>Kappa, unspecified</td>
<td>0.53 (0.53 – 0.54)</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiig et al.</td>
<td>63-158</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.56 – 0.70</td>
<td>Moderate to good agreement</td>
</tr>
<tr>
<td>(2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pietrzak et al.</td>
<td>63</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.65 (0.61 – 0.70)</td>
<td>Good agreement</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sambandam et al.</td>
<td>44</td>
<td>2</td>
<td>Weighted kappa</td>
<td>0.72</td>
<td>Good agreement</td>
</tr>
<tr>
<td>(2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n is number of patients; Obs is number of observers, * Interpretation of kappa values (Landis and Koch 1977): < 0.2 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 good agreement

It shows inconsistent levels of interobserver agreement for both classification systems, however the overall impression is that the majority of papers concluded with moderate levels of interobserver agreement (in 5 of 7 studies assessing the Catterall classification and in 5 of 7 studies assessing the original lateral pillar classification). The recent modification of the Herring’s lateral pillar classification was initially reported to be as reliable as its original version (Herring et al. 2004). The introduction of the borderline B/C group led to decreased interobserver agreement in paper IV, which is in accordance with the one study that has
been performed independently from the Herring group (Rajan et al. 2013). The authors showed fair interobserver agreement after the introduction of the B/C borderline group. The Herring group (Herring et al. 2004) found in average good interobserver agreement for the modified lateral pillar classification (kappa = 0.71) using a quadratic weighted kappa analysis. The authors attributed only half as much importance in misclassifying the borderline group B/C to its adjacent groups as they attributed misclassifying between groups A, B and C. Since quadratic weighting may increase kappa values artificially if an extra category is introduced within a classification system (Brenner and Kliebsch 1996), this may have led to an unrepresentatively high level of agreement.

7.3.2.3 Modified Stulberg classification at 5-year follow-up

The radiographic outcome in paper II and III has been classified at 5-year follow-up according to the modified Stulberg´s classification. In paper IV, we retested the interobserver agreement of the modified Stulberg classification and found lower levels of interobserver agreement (ranging from fair to moderate), than reported by our group previously (Wiig et al. 2007). We therefore felt the need to reassess the reproducibility of modified Stulberg classification by a multi-observer approach. We determined the modified Stulberg category, which has been assigned by the majority of the 4 observers for each radiograph. We termed the category as “true” if there was more than 50% consensus among the observers. All observers agreed upon the hips with a flat femoral head shape. Disagreement was found in 6 of the remaining 30 cases where observers could not agreed on whether the shape was round or ovoid. These radiographs were reassessed by 2 observers. Loss of height within 2 mm of a concentric circle on AP and frog-leg projection was defined as round and more than 2 mm as ovoid.
7.3.3 Prognostic factors for radiographic outcome

Paper II-IV investigated whether acetabular changes, bilateral hip affection and the extent of femoral head necrosis at an early stage may be associated with the femoral head sphericity at 5-year follow-up.

7.3.3.1 Acetabular changes

The acetabulum in Perthes disease remained steep, with a wider and shallower shape compared to the non-affected hip (paper II). The latter observation was especially marked in the early phase (initial and fragmentation stage) and these changes have been associated with a poorer outcome (Bellemans et al. 1979). Our results could not confirm this association, but we rather found a tendency that a wider and shallower acetabulum would be accompanied of a more spherical femoral head form (Paper II). It is well understood that the size of the femoral head in Perthes disease often exceed that of the unaffected side. We hypothesized that acetabular adaptation to the coxa magna would promote accommodation of the enlarged femoral head. A wider and shallower acetabulum would contribute to hip joint congruency and containment, maintaining the sphericity of the femoral head in the long term.

The importance of the lateral acetabular margin has been examined in DDH and Grzegorzewski et al. (2006) classified hips according to the slope of the lateral margin in Perthes disease. They found an association between an upward-sloping lateral margin and a flat femoral head shape at skeletal maturity. They hypothesized that a deformed lateral margin could lead to decreased anatomical support for the femoral head and, thereby, cause further subluxation and flattening. Our results do not support these findings, as we found no significant association between the lateral acetabular inclination early in the course and femoral head shape 5 years after diagnosis.
7.3.3.2 Bilateral involvement

In total, there were 55 patients (13%) with bilateral hip affection (paper III). For
investigation of the natural course we assessed all non-operatively treated patients (n=40).
We could identify 2 groups with different mode of onset; one group (n=23) where hips were
simultaneously affected and a second group (n=17) where hips developed Perthes disease
sequentially. In the latter group, the risk of developing femoral head in the second hip was 5
%, which is higher than the prevalence in a general population (Wii et al. 2006). The mean
age of onset was similar in both groups (5.4 years and 5.5 years respectively). In children
with sequential onset, the second hip was diagnosed with a mean of 1.9 years after the first
one, which is in accordance with other authors (Van den Bogaert et al. 1999). We assessed
whether the onset pattern would have an impact on the course of the disease and thereby
influence the final femoral head shape. Sequential onset was associated with a higher
degree of femoral head deformity whereas concurrent hip onset gave similar outcome as in
hips with unilateral affection.

The age at onset has been established as one of the major prognostic factors in terms of
the late radiographic outcome. One may argue that worse hip deformity as seen in the
second hip in sequential onset affection, may be due to the later affection. We implemented
age at diagnosis for each hip as a cofounding factor in the analysis in order to eradicate for
this significant bias. We classified the bilateral hips according to the modified Catterall
classification (2 groups) and found that 76% of the hips had more than 50% of femoral head
necrosis. Van den Bogaert et al. found similar high percentage of Catterall 3 and 4 hips
among the bilateral cases (Van den Bogaert et al. 1999). The extent of femoral head necrosis
was also similar compared to the natural history of unilateral cases assessed by our study
group (Terjesen et al. 2010). However, we adjusted the multivariable regression analysis for
the extent of femoral head necrosis since femoral head involvement has been shown to be of importance for the long-term prognosis (Herring et al. 2004, Wiig et al. 2008, Terjesen et al. 2010). Hence, sequential onset pattern had a significant bearing on the femoral head shape 5 years after diagnosis regardless of age at onset and extent of femoral head necrosis.

7.3.3.3 Extent of femoral head necrosis

In paper IV, we investigated the associations between 3 classification systems at fragmentation and the femoral head sphericity at 5-year follow-up. The original 3-group lateral pillar classification had the strongest association, which is in accordance with previous studies (Herring et al. 1992, Ritterbusch et al. 1993, Farsetti et al. 1995, Ismail and Macnicol 1998, Lappin et al. 2002, Terjesen et al. 2010). All hips that had a flat femoral head shape at 5-year follow-up in our study had been classified as group C at the fragmentation phase. The Catterall classification as predictor of radiographic outcome has been more controversial. It did not correlate well with the final radiographic outcome in some studies (Weinstein 1985, Ismail and Macnicol 1998, Gigante et al. 2002), while others confirmed it as a prognostic factor (Dickens and Menelaus 1978, Meurer et al. 1999, Terjesen et al. 2010). Especially when modified into a 2-group classification, distinguishing between more and less than 50% of femoral head necrosis, it was a strong predictor of radiographic outcome (Wiig et al. 2008). Our results confirm a significant association between the Catterall 4-group classification and the femoral head sphericity at 5-year follow-up, but the association seemed to be somewhat weaker than that of the original lateral pillar classification.

The Herring group reviewed the hips in their original study and identified a group of hips with radiographic findings more severe than those typical of group B but less severe than those in group C (Herring et al. 2004). The authors introduced a new classifications group (borderline B/C) and stated that the modified classification was a strong prognostic factor for
the final radiographic outcome. We found moderate association between the modified 4-group lateral pillar classification and modified Stulberg as outcome variable, while the original 3-group lateral pillar classification was strongly associated. The findings of other authors (Froberg et al. 2011) confirmed a somewhat weaker association of the modified lateral pillar classification.
8 Conclusions

1.) We observed a pronounced initial flattening and widening of the acetabulum which partly remodeled 5 years after diagnosis. The lateral acetabular margin failed to evolve normally and the inclination of the acetabulum remained steep, not following the natural hip development.

2.) Sharp’s angle and the acetabular depth- width ratio (ADR) showed good interobserver agreement and reliability and they can therefore be recommended in clinical practice.
Acetabular retroversion measured by ISS showed inconclusive agreement and the observations for lateral acetabular inclination (LAI) had fair agreement. Thus, they can hardly be recommended in clinical practice.

3.) Early widening and flattening of the acetabulum was not associated with a flat femoral head 5 years after diagnosis. There was a tendency that children with a wider and shallower acetabulum had a lesser degree of flattening of the femoral head 5 years after diagnosis.

4.) Bilateral Perthes disease occurred in 13 % of our study population. Children with concurrent onset of bilateral disease had similar radiographic outcome as children with unilateral involvement, whereas children with sequential onset had worse radiographic outcome. Following a diagnosis of Perthes disease in one hip, there was a 5% chance of developing a femoral head necrosis in the contralateral hip.
5.) The Catterall classification and the original lateral pillar classification had moderate interobserver agreement, while the modified lateral pillar classification had fair agreement. The strongest association between the extent of femoral head necrosis at an early stage of the disease and the femoral head shape at 5-year follow-up was found for the original lateral pillar classification, followed by the Catteral classification. The introduction of the borderline B/C group did not increase the agreement or the prognostic value of the lateral pillar classification, which raises concerns about its usefulness in clinical practice.
9 Future Research

Sophisticated MRI techniques, such as gadolinium-enhanced perfusion MRI, are able to detect the intra-cartilaginous vascular structures of the growing skeleton. Hence, imaging of the vascularity and revascularization processes in Perthes disease could be obtained and understood more precisely (Kim et al. 2016). In a first report, MRI perfusion images of the epiphysis detected the extent of femoral head necrosis significantly earlier than conventional radiography (Kim et al. 2014). Future research will have to investigate whether these findings are reproducible and eligible for a new systematic MRI based classification. Perthes treatment protocols are currently initiated according to the age of onset and femoral head involvement as seen on conventional radiographs. In many cases, deformity of the femoral head has occurred at the time of radiographic classification. The recent MRI advances with significant earlier detection of severe Perthes hip affection may be of great potential since treatment can be initiated prior to femoral head deformity. Future research will have to clarify whether early treatment based on early MRI detection has the potential to prevent femoral head deformity. MRI has further shown to depict the deformities more precisely at all stages of Perthes disease, also reformatted in a 3 D model (Standefer et al. 2016). Residual hip deformity is currently evaluated based on the femoral head shape as seen on 2 projections of conventional radiographs. Future research has to evaluate whether MRI based appraisal of the final radiographic outcome may present an advantage compared to the conventional classifications.
10 References


The acetabulum in Perthes’ disease: Inter-observer agreement and reliability of radiographic measurements

Stefan Huhnstock¹, Svein Svenningsen², Are H Pripp³, Terje Terjesen¹, and Ola Wiig¹

¹Orthopaedic Department, Oslo University Hospital, Oslo; ²Orthopaedic Department, Sørlandet Hospital, Arendal; ³Department of Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway.
Correspondence: Stefan.Huhnstock@oslo-universitetssykehus.no
Submitted 13-11-22. Accepted 14-04-15

Background and purpose — Perthes’ disease leads to radiographic changes in both the femoral head and the acetabulum. We investigated the inter-observer agreement and reliability of 4 radiographic measurements assessing the acetabular changes.

Patients and methods — We included 123 children with unilateral involvement, femoral head necrosis of more than 50%, and age at diagnosis of 6 years or older. Radiographs were taken at onset, and 1 year and 5 years after diagnosis. Sharp’s angle, acetabular depth-width ratio (ADR), lateral acetabular inclination (LAI), and acetabular retroversion (ischial spine sign, ISS) were measured by 3 observers. Before measuring, 2 of the observers had a consensus meeting.

Results — We found good agreement and moderate to excellent reliability for Sharp’s angle for all observers (intra-class correlation coefficient (ICC) > 0.80 with consensus, ICC = 0.46–0.57 without consensus). There was good agreement and substantial reliability for ADR between the observers who had had a consensus meeting (ICC = 0.62–0.89). Low levels of agreement and poor reliability were found for observers who had not had a consensus meeting. LAI showed fair agreement throughout the course of the disease (kappa = 0.28–0.52). The agreement between observations for ISS ranged from fair to good (kappa = 0.20–0.76).

Interpretation — Sharp’s angle showed the highest reliability and agreement throughout the course of the disease. ADR was only reliable and showed good agreement between the observers when landmarks were clarified before measuring the radiographs. Thus, we recommend both parameters in clinical practice, provided a consensus is established for ADR. The observations for LAI had only fair agreement and ISS showed inconclusive agreement in our study. Thus, LAI and ISS can hardly be recommended in clinical practice.

Perthes’ disease leads to typical radiographic changes of the femoral head. Several authors have described simultaneous changes of the acetabular anatomy, such as hypertrophy, bicompartamental development, retroversion, and dysplastic changes (Yngve and Roberts 1985, Joseph 1989, Ezoe et al. 2006).

Most measurements describing the radiographic changes of the acetabulum on anteroposterior (AP) pelvic radiographs have been validated in children with hip dysplasia. As the hip pathology and the morphological changes in Perthes’ disease are different from those in hip dysplasia, we wanted to assess inter-observer reliability and agreement of 4 commonly used acetabular measurements at the different stages of skeletal maturity in Perthes’ disease.

Patients and methods

In the Norwegian prospective multicenter study on Perthes’ disease, 425 patients were registered between 1996 and 2000 (Wiig et al. 2008). Radiographs were taken at onset, and at 1 and 5 years after diagnosis. Based on AP and Lauenstein projections, the affected hips were classified according to the original Catterall classification (1971). For the present study, we included all patients with more than 50% femoral head necrosis (groups 3 and 4), unilateral involvement, and age at onset of 6 years or older (n = 152). We analyzed affected and unaffected hips only if acetabular landmarks were adequately exposed. Thus, another 29 children had to be excluded. The mean age at time of diagnosis of the remaining 123 cases was 7.5 years (SD 1.2) (90 boys and 33 girls).

4 different radiographic parameters were measured on AP pelvic radiographs to assess the acetabular anatomy:

- **Sharp’s angle**

  This angle was described by Sharp (1961) in the assessment of hip dysplasia. A reference line was drawn between the inferior points of the teardrops on AP pelvic radiographs. The angle was formed by this reference line and a line connecting the inferior point of the teardrop and lateral edge of the acetabu-
The angle was measured in the affected and the unaffected hip.

**Acetabular depth-width ratio (ADR)**

The acetabular depth and width were measured on AP pelvic radiographs as described by Heyman and Herndon (1950). The width was defined as a line connecting the upper osseous acetabular margin and the lower end of the teardrop. This landmark is often more accurately defined than the lower acetabular margin. The depth was defined as the distance from the width line to the deepest point of the acetabulum (Figure 1).

For this measurement, 1 additional patient had to be excluded because of unsatisfactory exposure of the fossa acetabuli, due to a radiographic shielding device. We defined the ADR according to Cooperman et al. (1983) as 
\[ \text{ADR} = \left( \frac{\text{depth}}{\text{width}} \right) \times 1,000. \]

**Lateral acetabular inclination**

The lateral acetabular inclination was introduced by Cooperman et al. (1983) and later applied by Grzegorzewski et al. (2006) to children with Perthes’ disease. It was defined as down, horizontal, or up depending on whether the lateral lip of the acetabulum was below the weight-bearing dome of the acetabulum, horizontal, or above the weight-bearing dome of the acetabulum (Figure 2).

**Acetabular retroversion**

When the ischial spine is visible inside the pelvic inlet on a standardized AP pelvic radiograph, there is a prominence of the ischial spine. This may indicate acetabular retroversion. This sign has been suggested as an alternative measurement of acetabular retroversion to the more commonly used crossover sign in skeletally immature patients (Kalberer et al. 2008). We considered the ischial spine sign (ISS) to be positive if the ischial spine protruded beyond the pelvic rim into the pelvic inlet on standardized AP radiographs.

We considered radiographs to be standardized if they met the criteria for symmetric pelvic rotation as outlined by Siebenrock et al. (2003). These are symmetric appearance of the obturator foramina and the tip of the coccyx pointing toward the symphysis pubis.

The measurements were performed manually by 3 observers using a standardized goniometer. None of the radiographs contained any informative landmarks. All measurements were performed independently.

Observer 1 (SH): A resident in orthopedic surgery with special interest in pediatric orthopedic surgery. He assessed all radiographs at onset, and 1 and 5 years after diagnosis (n = 369) (Figure 3). The observer was briefed on the theoretical basis and practical use of the radiographic parameters in a consensus-building meeting by a consultant in pediatric orthopedic surgery (OW) before measuring the radiographs.

Observer 2 (OW): A consultant in pediatric orthopedic surgery. He measured the AP pelvic radiographs of 57 patients at the time of diagnosis, and at 1 and 5-year follow-up (n = 171). The radiographs of every other patient (alphabetically) were selected (total n = 61). The radiographic films of 4 patients from 3 local hospitals could not be retrieved for the assessment by observer 2; thus, 57 patients were examined.

Observer 3 (SS): A consultant in orthopedic surgery with great experience in examining radiographs of hips in children. He assessed Sharp’s angle (n = 123) and acetabular depth and width (n =122) in radiographs taken 5 years after diagnosis.

**Statistics**

Several statistical strategies have been described in the evaluation of reproducibility in measurement studies for numerical data (Bland and Altman 1986, 1999, Petrie 2006). The term reproducibility includes both agreement and reliability, and these 2 terms are often used interchangeably (Guyatt et al. 1987, Stratford and Goldsmith 1997, de Vet et al. 2006, Lee et al. 2012). Reliability parameters (e.g. intra-class correlation
coefficients (ICCs)) are related to how well measurements can be distinguished from each other, while agreement parameters, as used in the Bland-Altman method (1986), assess how close scores for repeated measurements are. We analyzed ICC using a 1-way random-effect model assuming a single measurement (McGraw and Wong 1996). An ICC of 0 indicates no more reliability than would be expected by chance alone, whereas values close to 1 indicate perfect reliability. We interpreted the intermediate values according to Landis and Koch (1977): values of less than 0.01 indicate poor reliability; 0.01 to 0.20, slight reliability; 0.21 to 0.40, fair reliability; 0.41 to 0.60, moderate reliability; 0.61 to 0.80, substantial reliability; and more than 0.80, excellent reliability.

We used the Bland-Altman method to examine the differences in numerical data between observers (Bland and Altman 1986, 1999). We calculated the differences between observations of 2 observers for each individual and calculated the mean and the standard deviation of the difference distribution. We defined good agreement to be when mean differences between the observers were less than 5% of their respective mean values. The 95% limits of agreement were calculated as the mean difference between the 2 measurements ± 1.96 SD. This range includes 95% of the interobserver differences.

The categorical data were analyzed with kappa statistics (Cohen 1968). For analysis with 3 or more selected categories, kappa statistics with linear weighting was used, defining the imputed relative distances between ordinal categories as 1. Like the ICC for continuous data, kappa is a measure of agreement between 2 sets of categorical data (Fleiss and Cohen 1973). Kappa has a maximum of 1 when agreement is perfect and a value of 0 indicates agreement no better than chance. As suggested by Altmann (1999), we interpreted the kappa values as follows: values of less than 0.20 indicate poor agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, good agreement; and greater than 0.80, very good agreement. The multirater kappa statistics are commonly used to describe chance-corrected agreement (Landis and Koch 1977, Posner et al. 1990, McHugh 2012). Statistical analysis was done using SPSS software version 20.

**Ethics**
Recruitment of patients was done by obtaining informed consent, and the study was approved by the Norwegian Data Inspectorate and the Norwegian Directorate of Health and Social Affairs in 1995.

**Results**

**Sharp’s angle**
As measured by observer SH, the mean value of Sharp’s angle at the time of diagnosis was 45° for the affected hip (Table 1).
and it remained stationary during follow-up. The mean angle decreased statistically significantly in normal hips from 45° at diagnosis to 42° at 5-year follow-up (p < 0.01) (Table 1).

Observers SH and OW assessed radiographs of 57 patients at the time of diagnosis and at 1- and 5-year follow-up. There were low inter-observer differences between each pair of observations for radiographs taken at diagnosis and at 1- and 5-year follow-up, indicating good agreement in both the affected and unaffected hips (Table 1). The range, which included 95% of the inter-observer differences, was narrow and showed negligible differences between radiographs taken at diagnosis and at 1- and 5-year follow-up, and between normal and affected hips. Excellent inter-observer reliability was found for the affected hips (ICC > 0.80), whereas substantial to excellent agreement was noted for normal hips (ICC = 0.65–0.88).

Observers SH and SS measured all radiographs at the 5-year follow-up (n = 123). The mean value for the affected hip was 46° for SH and 44° for SS (Table 1), whereas lower mean values were found for the unaffected hip, at 42° (SH) and 44° (SS). Using the Bland-Altman method, we found that mean differences between the observers were below 5% of their mean values, indicating good inter-observer agreement. For both the affected hip and the unaffected hip, moderate reliability was found with ICC values ranging from 0.52 to 0.57 (Table 1).

Observers SH and OW assessed radiographs of 57 patients at the time of diagnosis and at 1- and 5-year follow-up. There were low inter-observer differences between each pair of observations for radiographs taken at diagnosis and at 1- and 5-year follow-up, indicating good agreement in both the affected and unaffected hips (Table 1). The range, which included 95% of the inter-observer differences, was narrow and showed negligible differences between radiographs taken at diagnosis and at 1- and 5-year follow-up, and between normal and affected hips. Excellent inter-observer reliability was found for the affected hips (ICC > 0.80), whereas substantial to excellent agreement was noted for normal hips (ICC = 0.65–0.88).

Acetabular depth-width ratio (ADR)

Observer SH found a mean ADR of 284 in affected hips at the time of diagnosis and significantly lower ADR values 1 and 5 years after diagnosis (262 and 263) (Table 2). We observed higher ADR values for the unaffected hip, and they remained unchanged throughout the course of the disease.

Low inter-observer differences between SH and OW were noted for affected hips and unaffected hips (Table 2). The 95% limits of agreement were wide (Figure 4), indicating poor inter-observer agreement. Similarly, we found only fair reliability between the observers, with ICC = 0.31 for affected hips and ICC = 0.23 for unaffected hips (Table 2).

Observers SH and SS assessed 122 radiographs taken at the 5-year follow-up (Table 2). The mean differences between the observers exceeded 10% of their mean values and the 95% limits of agreement were rather wide, indicating poor inter-observer agreement. Similarly, we found only fair reliability between the observers, with ICC = 0.31 for affected hips and ICC = 0.23 for unaffected hips (Table 2).

Inter-observer differences between observers OW and SS (n = 56) were higher for both affected and unaffected hips. The 95% limits of agreement showed wide measurement distribution, indicating lower levels of agreement. Poor reliability was found for unaffected hips (ICC = 0.05) and fair reliability was found for affected hips (ICC = 0.37) (Table 2).

Lateral acetabular inclination

We found fair to moderate agreement between observers SH and OW (n = 57) for the affected hip, with kappa values increasing slightly from 0.40 at the time of diagnosis to 0.46 at the 5-year follow-up (Table 3). Similarly, we obtained fair to
moderate agreement in the unaffected hips, with kappa values ranging from 0.20 to 0.65.

**Acetabular retroversion**

29 radiographs that met the criteria for symmetric pelvic rotation at the time of diagnosis were assessed. We found moderate agreement between observers SH and OW for the affected hips (kappa = 0.52, CI: 0.22–0.82) and fair agreement for unaffected hips (kappa = 0.20, CI: 0–0.65). At the 1-year follow-up, 17 radiographs met the criteria for symmetric rotation. We found good inter-observer agreement for the affected side (kappa = 0.76, CI: 0.46–1), whereas fair agreement was obtained for unaffected hips (kappa = 0.26, CI: 0–0.89). At the 5-year follow-up, 20 radiographs met the criteria for symmetric rotation and we found good inter-observer agreement for

---

**Table 2. Inter-observer measurements of ADR**

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Mean difference (SD)</th>
<th>95% limits of agreement</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SH</td>
<td>OW</td>
<td>SS</td>
<td>SH-OW</td>
</tr>
<tr>
<td><strong>Perthes hip</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At diagnosis</td>
<td>284</td>
<td>295</td>
<td>–10</td>
<td>–76 to 57</td>
</tr>
<tr>
<td>(n = 56)</td>
<td>(40)</td>
<td>(46)</td>
<td>(34)</td>
<td></td>
</tr>
<tr>
<td>1-year follow-up</td>
<td>262</td>
<td>272</td>
<td>–9</td>
<td>–100 to 83</td>
</tr>
<tr>
<td>(n = 56)</td>
<td>(47)</td>
<td>(61)</td>
<td>(47)</td>
<td></td>
</tr>
<tr>
<td>5-year follow-up</td>
<td>263</td>
<td>267</td>
<td>292</td>
<td>–4</td>
</tr>
<tr>
<td>(n = 122)</td>
<td>(43)</td>
<td>(46)</td>
<td>(46)</td>
<td>(21)</td>
</tr>
<tr>
<td><strong>Unaffected hip</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At diagnosis</td>
<td>306</td>
<td>312</td>
<td>–7</td>
<td>–55 to 40</td>
</tr>
<tr>
<td>(n = 56)</td>
<td>(25)</td>
<td>(38)</td>
<td>(24)</td>
<td></td>
</tr>
<tr>
<td>1-year follow-up</td>
<td>310</td>
<td>329</td>
<td>–17</td>
<td>–80 to 47</td>
</tr>
<tr>
<td>(n = 56)</td>
<td>(32)</td>
<td>(43)</td>
<td>(32)</td>
<td></td>
</tr>
<tr>
<td>5-year follow-up</td>
<td>309</td>
<td>323</td>
<td>307</td>
<td>–14</td>
</tr>
<tr>
<td>(n = 122)</td>
<td>(38)</td>
<td>(46)</td>
<td>(197)</td>
<td>(27)</td>
</tr>
</tbody>
</table>

n: number of patients; ICC: intra-class correlation coefficient.

---

**Table 3. Inter-observer agreement for lateral acetabular inclination**

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perthes hips</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At diagnosis</td>
<td>0.40</td>
<td>0.10</td>
<td>0.21–0.59</td>
</tr>
<tr>
<td>1-year follow-up</td>
<td>0.45</td>
<td>0.10</td>
<td>0.26–0.63</td>
</tr>
<tr>
<td>5-year follow-up</td>
<td>0.46</td>
<td>0.10</td>
<td>0.22–0.60</td>
</tr>
<tr>
<td><strong>Unaffected hips</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At diagnosis</td>
<td>0.38</td>
<td>0.11</td>
<td>0.16–0.60</td>
</tr>
<tr>
<td>1-year follow-up</td>
<td>0.51</td>
<td>0.12</td>
<td>0.27–0.75</td>
</tr>
<tr>
<td>5-year follow-up</td>
<td>0.28</td>
<td>0.13</td>
<td>0.01–0.54</td>
</tr>
</tbody>
</table>

Kappa: kappa analysis with linear weighting.
the affected side (kappa = 0.79, CI: 0.53–1). We found moderate agreement for the normal hips (kappa = 0.49, CI: 0–0.98).

Discussion
In order to describe the acetabular changes in Perthes’ disease properly, there is a need for reliable radiographic measurements that should be easy to use, have good inter- and intra-rater agreement, and have prognostic value.

As part of the Norwegian national prospective study on Perthes’ disease, we have assessed the inter-observer agreement and reliability of 4 commonly used acetabular measurements in children with age at disease onset of 6 years or more, and more than 50% femoral head necrosis (Van den Bogaert et al. 1999). This multicenter study involved 28 hospitals throughout Norway and we were not able to standardize the radiographs, which is an obvious limitation of the study.

Radiographic classifications in Perthes’ disease have been subject to validation in previous studies (Mahadeva et al. 2010); however, only a few authors have reported on inter-observer agreement and reliability of radiographic measurements in this condition (Wiig et al. 2002).

Sharp’s angle
To our knowledge, Sharp’s angle has never been validated in children with Perthes’ disease. Nelitz et al. (1999) reported substantial inter-observer reliability for Sharp’s angle (ICC = 0.74–0.78) in skeletally mature patients with DDH. These results were similar to those of Engesæter et al. (2012), who obtained excellent inter-observer reliability (ICC = 0.83) in 18- to 19-year-old healthy women. Furthermore, Engesæter et al. reported mean differences for each pair of observations of between 2.0% and 7.2% of the mean values and narrow 95% limits of agreement, indicating good agreement. Agus et al. (2002) found that Sharp’s angle measurement was a reliable measurement in skeletally immature children with DDH (mean age 9.5 years). Our inter-observer findings are in accordance with those of previous authors, as we could demonstrate good inter-observer reliability as well as good agreement.

Acetabular depth-width ratio (ADR)
Heyman and Herndon (1950) showed that acetabular width and depth were altered in Perthes’ disease. They defined that the acetabular depth-width ratio is one of the major criteria describing characteristic radiological changes. To our knowledge, no validation of this parameter in Perthes’ disease has been published. However, some authors have assessed the inter-observer reliability and agreement in children with DDH, but the findings differed widely. Takatori et al. (2010) reported large values for the coefficient of variation for ADR, indicating a low degree of agreement, which was in accordance with the results of Clohisy et al. (2009) who demonstrated fair reliability for acetabular depth in adults. In contrast, Nelitz et al. (1999) found moderate inter-observer reliability for ADR with ICC values ranging from 0.58 to 0.63. Engesæter et al. (2012) showed substantial inter-observer reliability for ADR (ICC = 0.77) in patients with DDH. Both publications reported good inter-observer agreement.

In the present study, observers who had a consensus-building meeting before performing the measurements had low differences between each pair of observations and a narrow range of 95% limits of agreement, indicating good inter-observer agreement. ICC values ranged from 0.62 to 0.89, indicating substantial to excellent reliability. In contrast to this, we found a wide range for the 95% limits of agreement and higher differences between the observations for observers without a consensus-building meeting, indicating poorer inter-observer agreement. ICC values for non-consensus observers ranged from 0.07 to 0.37, indicating poor to fair reliability of ADR.

Lateral acetabular inclination
To our knowledge, no previous study has assessed the inter-observer reliability and agreement of this parameter. We found fair to moderate agreement between observers (kappa = 0.28–0.51) in both affected and unaffected hips regardless of when the measurements were performed during the course of the disease. Based on our findings, acetabular inclination appears to be less suitable in radiographic evaluation of Perthes’ disease.

Acetabular retroversion
Assessment of the anatomy of a 3-dimensional object such as the acetabulum on a 2-dimensional radiograph has obvious limitations. However, the crossover sign (Jamali et al. 2007) and the ISS (Kalberer et al. 2008) revealed acetabular retroversion in AP pelvic radiographs. Of the 57 radiographs that were available to inter-observer analysis, 28 radiographs had to be excluded at time of diagnosis, 40 at 1-year follow-up, and 37 at 5-year follow-up due to radiographic standardization criteria. Perthes’ disease leads to muscular atrophy and contracture of the affected side, which may cause pelvic rotation along the longitudinal axis on AP pelvic radiographs (Schiller and Axer 1972). This might be the reason for the high number of dropouts in our study population. Other authors who assessed acetabular retroversion in Perthes’ disease had a similar reduction of radiographs included (Larson et al. 2011, Kawahara et al. 2012). To our knowledge, only 1 study so far has assessed the inter-observer agreement of acetabular retroversion using ISS (Kappe et al. 2011). The authors examined AP pelvic radiographs of 20 skeletally mature patients and the number of patients included was similar to the number of radiographs assessed at diagnosis, and at 1-year and 5-year follow-up in our study. Kappe et al. (2011) emphasized that acetabular version was subject to considerable inter-observer differences and related to the individual experience of observers. Our findings are in accordance with these results, as we had moderate to good inter-observer agreement for affected hips but only fair to moderate inter-observer agreement for unaffected hips. This
indicates that acetabular retroversion as assessed on pelvic radiographs is hardly a suitable parameter in Perthes’ disease.

In conclusion, Sharp’s angle had the highest inter-observer reliability and agreement of the 4 parameters investigated—throughout the course of Perthes’ disease in skeletally immature patients. ADR was reliable and showed good agreement between the observers only when landmarks were clarified before measuring the radiographs. Thus, we recommend both parameters in clinical practice, provided a common consensus is established and that these parameters are of prognostic value, which will be evaluated in a separate study. The observations for lateral inclination of the acetabulum had only fair to moderate agreement and ISS showed inconclusive agreement in our study. Thus, these parameters can hardly be recommended in clinical practice.

No competing interests declared.


Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70 (4): 213-20.


The acetabulum in Perthes’ disease: a prospective study of 123 children

S. Huhnstock · S. Svenningsen · A. H. Pripp · T. Terjesen · O. Wiig

Abstract

Purpose We assessed the radiographic changes of the acetabulum during the course of Perthes’ disease and investigated whether they were associated with femoral head sphericity 5 years after diagnosis.

Methods We studied 123 children with unilateral Perthes’ disease, femoral head necrosis more than 50 % and age at diagnosis 6 years or older. Pelvic radiographs were taken at onset, 1 year and 5 years after diagnosis. Sharp’s angle, acetabular depth-to-width ratio (ADR) and lateral acetabular inclination were measured.

Results Compared to the unaffected hips, the Perthes’ hips developed significantly higher Sharp’s angles ($p < 0.001$) and a higher proportion with an upward-sloping lateral acetabular margin (Perthes’ hips: 49 %, unaffected hips 1 %). The mean ADR values were significantly lower on the affected side at all stages ($p < 0.001$). ADR values at diagnosis were associated with a more spherical femoral head at the 5-year follow-up [odds ratio (OR) 1.012, 95 % confidence interval (CI) 1.002–1.022, $p = 0.016$]. None of the other acetabular parameters were significantly associated with the femoral head shape 5 years after diagnosis.

Conclusion The acetabulum developed an increasingly dysplastic shape in the course of Perthes’ disease. Early dysplastic changes of the acetabulum were not associated with a poor radiological outcome 5 years after diagnosis. Routine measurement and monitoring of acetabular changes in plain radiographs were of little prognostic value and can, therefore, hardly be recommended in clinical practice.

Keywords Legg–Calvé–Perthes’ disease · Acetabulum · Hip · Secondary hip dysplasia

Introduction

Perthes’ disease leads to typical anatomic changes of the femoral head [1–3]. In 1950, Heyman and Herndon emphasised that radiographic alteration of the acetabulum plays a crucial role in the evolution of the disease and they defined it as one of four major radiographic criteria to describe this condition [4]. In the last few decades, several authors have reported on radiographic changes of the acetabular anatomy, such as hypertrophy, retroversion, bicompartamental and dysplastic development [5–7]. It is still unknown as to whether these changes occur primarily or if they are secondary to the anatomical changes of the femoral head. The aim of our study was to describe radiographic changes of the acetabulum during the course of the disease and to assess whether early acetabular changes were associated with femoral head sphericity 5 years after diagnosis.

Patients and methods

As part of the Norwegian prospective multi-centre study on Perthes’ disease, 425 patients were registered between
We analysed both the affected as well as the unaffected hips in all cases with unilateral involvement, age at onset 6 years or older and femoral head necrosis more than 50% (n = 152). Radiographs were taken at diagnosis and at 1- and 5-year follow-up. The degree of femoral head necrosis was assessed according to the original Catterall classification [9]. We included radiographs classified as groups III and IV. Twenty-nine children were excluded due to inadequate exposure of the acetabular landmarks. Thus, 123 children (90 boys and 33 girls) with a mean age at the time of diagnosis of 7.5 years (range 6–13 years) were studied.

The radiographic phase was determined at the time of diagnosis according to Waldenström [10]. Sixty-three hips were in the initial phase (51%), 48 were in the fragmentation phase (39%), five were in the reossification phase (4%) and seven hips had not been classified (6%).

We applied the original lateral pillar classification of Herring et al. [11] in 110 patients at the fragmentation phase. Sixty hips were classified as lateral pillar type B (54.5%) and 50 hips as lateral pillar type C (45.5%). The femoral head cover was calculated as the percentage of the femoral head medial to Perkin’s line compared to the width of the femoral head, both measured parallel to Hilgenreiner’s line [4].

The children included in this study received either physiotherapy (n = 55), Scottish Rite orthosis (n = 26) or proximal femoral varus osteotomy (n = 71) [8], according to the choice of the local orthopaedic surgeons. The decision was based on surgeons’ preferences, treatment philosophy and local tradition. We combined patients treated with physiotherapy and orthosis into a non-operative treatment group.

Three different radiographic measurements were examined on antero-posterior (AP) pelvic radiographs to assess the acetabular anatomy, described in the following sections.

**Sharp’s angle**

This angle was described by Sharp in 1961 for the assessment of hip dysplasia [12]. A reference line was drawn between the inferior points of the teardrops on AP pelvis radiographs. The angle was formed by this reference line and a line connecting the inferior point of the teardrop to the lateral edge of the acetabulum (Fig. 1).

**Acetabular depth-to-width ratio (ADR)**

The acetabular depth and width were measured on AP pelvic radiographs, as described by Heyman and Herndon in 1950 [4]. The length of a line connecting the lateral osseous acetabular margin and the lower end of the teardrop defined the width of the acetabulum. This teardrop is often more accurately defined than the lower acetabular margin. The depth was defined as the distance from the width line to the deepest point of the acetabulum (Fig. 1).

According to Cooperman et al. [13], we measured the ADR as depth/width × 1,000. We compared our results of
the affected and unaffected hips with the depth-to-width quotients of 600 skeletally immature normal hips published by Bellemans et al. [14]. They established a normal range of the depth-to-width quotient for girls and boys. For the present study, we used Bellemans et al.’s results for children aged 6 years and older (Fig. 5). Values below 2 standard deviations (SDs) of the mean depth-to-width range for 6-year-olds were classified as dysplastically altered (boys: ADR <261; girls: ADR <274). We chose an ADR cut-off value of 265 based on the mean ADR between boys and girls.

Lateral acetabular inclination

The lateral acetabular inclination was introduced by Cooperman et al. [13] in 1983 and later applied by Grzegorzewski et al. [15] for children with Perthes’ disease. It was recorded as down, horizontal or up, depending on whether the lateral lip of the acetabulum was below, horizontal or above the weight-bearing dome of the acetabulum, respectively (Fig. 2).

Femoral head sphericity

Five years after diagnosis, hips were classified according the modified Stulberg three-group classification [16], where group A hips (Stulberg groups I and II) have a spherical head, group B hips (Stulberg group III) have an ovoid femoral head and group C hips (Stulberg groups IV and V) have a flat femoral head.

Statistical analysis

We used IBM SPSS Statistics, version 20 for all statistical analyses. Numerical data were described as the mean and range. The mean differences between the affected and unaffected hips of patient groups were statistically compared with a paired-samples t-test. Categorical data were described using the number of observations (percentages) and analysed with cross-table analysis and Pearson’s Chi-squared tests.

An ordinal regression model was used to estimate the cumulative odds ratio (OR) for worse radiological outcome (modified Stulberg classification) with selected clinical predictor variables. The following variables were each analysed separately in an ordinal regression model: age at onset of the disease, gender, Catterall classification, lateral pillar classification, received treatment, femoral head coverage at 1-year follow-up and ADR. Further, those that were significantly associated with the Stulberg classification (with $p \leq 0.05$ as criteria) were included in the final multi-variable ordinal regression model to assess their predictive ability regarding worse radiological outcome. Results were regarded as statistically significant if the $p$-values were below 0.05.

Table 1 Sharp’s angle and ADR values at the time of diagnosis and 1- and 5-year follow-up

<table>
<thead>
<tr>
<th>Time of observation</th>
<th>Mean Sharp’s angle (SD)</th>
<th>Mean ADR (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perthes’ hip</td>
<td>Unaffected hip</td>
</tr>
<tr>
<td>At diagnosis</td>
<td>45.5 (3.8)</td>
<td>44.9 (3.6)</td>
</tr>
<tr>
<td>1-year follow-up</td>
<td>45.0 (4.4)</td>
<td>44.4 (3.7)</td>
</tr>
<tr>
<td>5-year follow-up</td>
<td>45.7 (4.4)</td>
<td>42.2 (3.7)</td>
</tr>
</tbody>
</table>

SD standard deviation, ADR acetabular depth-to-width ratio, $n$ number of patients; if the numbers are not specified in the table, it includes all 123 patients.

Results

Sharp’s angle

The mean value of the Sharp’s angle for the affected hip remained stationary throughout the course of the disease (Table 1). It decreased significantly in unaffected hips from diagnosis to the 5-year follow-up ($p < 0.001$). The
difference between the Sharp’s angles of the affected and unaffected hips at the 5-year follow-up was significant \((p < 0.001)\) (Table 1).

Acetabular depth-to-width ratio

The mean ADR value of the affected hips at the time of diagnosis was significantly lower compared with the unaffected side \((p < 0.001)\) (Table 1). This was due to both an increase of acetabular width \((p < 0.001)\) and a slight decrease of depth \((p < 0.001)\) (Fig. 3). We analysed the ADR at diagnosis separately according to each Waldenström radiographic phase. The mean ADR value of the affected hips in the initial phase was significantly lower compared with the unaffected side \((287 \text{ vs. } 302; p < 0.001)\).

The ADR of the affected hips decreased significantly from time of diagnosis to the 1-year follow-up \((p < 0.001)\) (Table 1), with both a significant increase of acetabular width \((p < 0.001)\) as well as depth \((p < 0.001)\) (Fig. 3). At the 5-year follow-up, the mean ADR had increased significantly from the 1-year follow-up \((p = 0.020)\). This partial normalisation was due to a stable acetabular width and a significant increase of the acetabular depth \((p = 0.003)\) (Fig. 3). However, the ADR values were significantly lower in the affected hips at the 1- and 5-year follow-up compared with those of the unaffected hips \((p < 0.001)\). In unaffected hips, the ADR did not change significantly during the course of the disease (Table 1).

Lateral acetabular inclination

The proportion of affected hips with the lateral lip of the acetabulum below the weight-bearing dome at the time of diagnosis \((53.0 \%)\) (Table 2). This proportion increased during the course of the disease to \(68.0 \%\) at the 5-year follow-up. At diagnosis and 1-year follow-up, none of the unaffected hips had the lateral lip above the weight-bearing dome. This was also the case 5 years after diagnosis, except for one hip.

Prognostic factors for a spherical femoral head at the 5-year follow-up

For the evaluation of factors that might influence the radiological outcome at the 5-year follow-up, we performed an ordinal regression analysis for a single variable for each of the following parameters: Sharp’s angle, lateral acetabular inclination, ADR, age at diagnosis, gender, Catterall classification, lateral pillar classification, treatment and femoral head coverage. The results are given in Table 3. Of the three acetabular measurements, only the ADR was significantly associated with the modified Stulberg classification at the 5-year follow-up. Lower ADR values at the time of diagnosis were associated with a spherical femoral head (group A) and higher ADR values with a flat femoral head (group C) (Fig. 4). Using the ADR cut-off value of 265 at the time of diagnosis, we grouped the affected hips into ‘normal’ \((ADRs >265, n = 84)\) and ‘wider-shallower hips’ \((ADRs <265, n = 39)\). Patients with normal acetabula at diagnosis developed a flat femoral head in \(34 \%\) of the cases, while patients with wider and shallower acetabula developed a flat femoral head in \(8 \%\) of the cases in the evolution of Perthes’ disease. In order to assess the predictive ability for femoral head sphericity at the 5-year follow-up, we performed a final multi-variable ordinal regression model including all parameters with significant association in the single-variable analysis. The results show that treatment, lateral pillar classification and Catterall classification were strongly associated with the femoral head sphericity (Table 4). The ADR at diagnosis showed a trend but the association was not statistically significant \((p = 0.061)\).
Discussion

Sharp’s angle

The unaffected acetabulum followed the normal age-dependent development of Sharp’s angle as reported by Ozc¸elik et al. [17], whereas the affected side remained steep at the 5-year-follow-up. One previous study has measured the Sharp’s angle in children with Perthes’ disease. Joseph showed that the Sharp’s angle of Perthes’ hips in male, skeletally immature patients was higher compared to the unaffected side (44.7° vs. 42.9°) [5]. Furthermore, he found a significant difference in the Sharp’s angle between affected and unaffected hips in symptomatic adult patients with healed Perthes’ disease (40.1° and 36.1°, respectively). These results are in accordance with our results, indicating that a steeper acetabulum is developed in Perthes’ hips during the course of the disease.

Acetabular depth-to-width ratio

Heyman and Herndon showed that acetabulum on the affected side developed increased width and decreased depth in children with Perthes’ disease [4]. The present study identified when these changes occurred during the course of the disease. We found an increase in acetabular width within the first year of the disease, but no further increase until the 5-year follow-up. After initial decrease, the acetabular depth normalised during the course of the disease, indicating that the acetabulum was remodelling to an anatomically more normal shape. However, affected acetabula were wider and shallower at all stages of the disease compared with the unaffected side.

Bellemans et al. [14] showed that the ADR values increased from a base level of approximately 250 at 3 years of age to values over 300 at the age of 8 years and older in 600 normal hips. We compared the ADR values of the affected side at time the of diagnosis with the ADR values reported by Bellemans et al. and found that a majority was below the normal range for both girls and boys (Fig. 5). This tendency was aggravated at the 1-year follow-up (Fig. 6). These findings indicate early dysplastic alteration of the acetabulum in terms of increased width and decreased depth, as reported by other authors [5, 14].

### Table 3
Simple ordinal regression analysis, i.e. separate analysis for each prognostic factor, with three-group Stulberg classification as the outcome variable

<table>
<thead>
<tr>
<th>Prognostic factor</th>
<th>OR</th>
<th>95 % CI</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral pillar classification</td>
<td>0.188</td>
<td>0.086–0.410</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Treatment</td>
<td>2.731</td>
<td>1.384–5.391</td>
<td>0.004</td>
</tr>
<tr>
<td>Catterall classification</td>
<td>0.156</td>
<td>0.073–0.332</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ADR at diagnosis</td>
<td>1.012</td>
<td>1.002–1.022</td>
<td>0.016</td>
</tr>
<tr>
<td>Femoral head coverage at 1 year</td>
<td>0.971</td>
<td>0.942–1.002</td>
<td>0.064</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>1.012</td>
<td>0.988–1.035</td>
<td>0.334</td>
</tr>
<tr>
<td>Gender</td>
<td>1.213</td>
<td>0.566–2.600</td>
<td>0.620</td>
</tr>
<tr>
<td>Sharp’s angle</td>
<td>1.021</td>
<td>0.036–1.114</td>
<td>0.641</td>
</tr>
<tr>
<td>Acetabular lateral inclination</td>
<td>0.792</td>
<td>0.219–2.858</td>
<td>0.721</td>
</tr>
<tr>
<td></td>
<td>1.558</td>
<td>0.731–3.323</td>
<td>0.251</td>
</tr>
</tbody>
</table>

*OR odds ratio, 95 % CI 95 % confidence interval, ADR acetabular depth-to-width ratio

* An ordinal regression analysis of a three-group categorical parameter with the Stulberg classification leads to two odds ratio results

### Table 4
Multi-variable ordinal regression analysis with the modified three-group Stulberg classification as the outcome variable

<table>
<thead>
<tr>
<th>Prognostic factor</th>
<th>OR</th>
<th>95 % CI</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral pillar classification</td>
<td>0.276</td>
<td>0.119–0.644</td>
<td>0.003</td>
</tr>
<tr>
<td>Treatment</td>
<td>2.927</td>
<td>1.349–6.350</td>
<td>0.007</td>
</tr>
<tr>
<td>Catterall classification</td>
<td>0.324</td>
<td>0.138–0.760</td>
<td>0.010</td>
</tr>
<tr>
<td>ADR at diagnosis</td>
<td>1.011</td>
<td>1.000–1.022</td>
<td>0.061</td>
</tr>
</tbody>
</table>

*OR odds ratio, 95 % CI 95 % confidence interval, ADR acetabular depth-to-width ratio

---

Fig. 4 Diagram showing the relation between the mean ADR mean values at the time of diagnosis and the modified Stulberg classification at the 5-year follow-up

Fig. 5 Diagram showing the relation between the mean ADR mean values at the time of diagnosis and the modified Stulberg classification at the 5-year follow-up
Lateral acetabular inclination

Grzegorzewski et al. [15] evaluated 243 cases of unilateral Perthes’ disease at fragmentation and at skeletal maturity. They found that the lateral acetabular margin of the affected hips was below the weight-bearing dome of the acetabulum in 32 %, horizontal in 56 % and above in 12 % at fragmentation. These results are in accordance with our observations for affected hips at the time of diagnosis (Table 2). There was a consistent increase of affected acetabula with an upward-sloping lateral margin until the 5-year follow-up, which is markedly higher than that reported by Grzegorzewski et al. [15]. However, direct comparison is difficult because of differences between the study populations.
Prognostic factors for a spherical femoral head

To our knowledge, no previous study has associated Sharp’s angle at an early stage of Perthes’ disease with the later shape of the femoral head. We found no significant associations between this parameter at the time of diagnosis and at the 1-year follow-up with the modified three-group Stulberg classification at the 5-year follow-up. Thus, Sharp’s angle is of no prognostic value in Perthes’ disease.

Grzegorzekski et al. [15] found an association between an upward-sloping lateral margin and a flat femoral head shape at skeletal maturity. They hypothesised that a deformed lateral margin could lead to decreased anatomical support for the femoral head and, thereby, cause further
subluxation and flattening of the femoral head. Our results do not support these findings, as we found no significant association between the lateral acetabular inclination early in the course and the femoral head shape 5 years after diagnosis. Therefore, routine measurement and monitoring of Sharp’s angle and lateral acetabular inclination can hardly be recommended in clinical practice.

As previous studies on this subject have shown, we found that lateral pillar height of more than 50 %, operative treatment and femoral head necrosis less than 75 % were the three strongest predictors for a spherical femoral head [8, 18]. Bellemans et al. [14] stated that a dysplastic acetabulum was associated with poor radiographic outcome in Perthes’ disease. This is not in accordance with our results, as we found that wider and shallower acetabula at diagnosis were associated with a more favourable late radiographic outcome. However, the ADR had no significant prognostic value in the multi-variable regression test and, therefore, has limited value in clinical practice.

Development over time and possible causes of acetabular changes in Perthes’ hips

Perthes’ disease leads to an enlargement and lateralisation of the femoral head, thereby altering the force transmission from the femoral head to the acetabulum. Madan et al. [19] showed significantly increased metabolic activity in the area of the triradiate cartilage of the affected hip at an early stage [5, 19]. They hypothesised that hyperaemia may lead to increased growth at the triradiate hip at an early stage. They stated that the alteration in acetabular growth was due to lateral pressure from the deforming femoral head. Our results support only to some extent the theory of mechanically induced alterations of the acetabulum, since a wider and shallower acetabulum was already present at diagnosis in all hips, including those that were in the initial radiographic phase. These changes were most likely not caused by mechanical forces alone.

Acetabular development occurs by a combination of enchondral, interstitial and appositional growth [20]. Our observations support the work of Joseph. He suggested that some of the early acetabular changes do not necessarily follow the alterations of the shape of the femoral head [5]. He and others showed significantly increased metabolic activity in the area of the triradiate cartilage of the affected hip at an early stage [5, 19]. They hypothesised that hyperaemia may lead to increased growth at the triradiate cartilage, causing a widening of the acetabulum. Growth alteration due to hyperaemia in close relation to the growth plate is a phenomenon also known in other conditions, such as juvenile idiopathic arthritis, where higher metabolic activity may cause excessive growth [21]. Several authors found a thickening of medial cartilage of the acetabulum due to inflammatory processes and increased metabolism [6, 22, 23], leading to hypertrophy and reduced acetabular depth, independent of mechanical forces. Joseph suggested that increased appositional growth was mediated by synovitis and hyperaemia, being a factor contributing to depth reduction [5].

In summary, our results suggest that the dysplastic changes of the acetabulum in Perthes’ disease are evoked by primary and secondary mechanisms. Primarily, the disease induces excessive growth, causing widening and hypertrophy of the acetabulum that might be induced by hyperaemia and higher metabolic activity. Secondarily, altered dimensions of the femoral head lead to mechanically induced changes, which inhibit the natural tilt of the acetabulum and cause an upward-sloping lateral margin. Early dysplastic changes of the acetabulum were not associated with a poor radiological outcome 5 years after diagnosis. Contrarily, there was a tendency that children with a wider and shallower acetabulum had a lesser degree of later femoral head deformation.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References
Radiographic classifications in Perthes disease
Interobserver agreement and association with femoral head sphericity at 5-year follow-up

Stefan HUHNSTOCK 1,5, Svein SVENNINGSEN 2, Else MERCKOLL 3, Anthony CATTERALL 4, Terje TERJESEN 1, and Ola WIIG 1

1 Department of Paediatric Orthopaedic Surgery, Oslo University Hospital, 2 Department of Orthopaedic Surgery, Sørlandet Hospital, Arendal, 3 Department of Radiology, Oslo University Hospital, Norway, 4 Royal National Orthopaedic Hospital, London, England, 5 Institute of Clinical Medicine, University of Oslo, Norway
Correspondence: stefan.huhnstock@oslo-universitetssykehus.no
Submitted 2017-02-21. Accepted 2017-05-12.

Background and purpose — Different radiographic classifications have been proposed for prediction of outcome in Perthes disease. We assessed whether the modified lateral pillar classification would provide more reliable interobserver agreement and prognostic value compared with the original lateral pillar classification and the Catterall classification.

Patients and methods — 42 patients (38 boys) with Perthes disease were included in the interobserver study. Their mean age at diagnosis was 6.5 (3–11) years. 5 observers classified the radiographs in 2 separate sessions according to the Catterall classification, the original and the modified lateral pillar classifications. Interobserver agreement was analysed using weighted kappa statistics. We assessed the associations between the classifications and femoral head sphericity at 5-year follow-up in 37 non-operatively treated patients in a crosstable analysis (Gamma statistics for ordinal variables, \( \gamma \)).

Results — The original lateral pillar and Catterall classifications showed moderate interobserver agreement (kappa 0.49 and 0.43, respectively) while the modified lateral pillar classification had fair agreement (kappa 0.40). The original lateral pillar classification was strongly associated with the 5-year radiographic outcome, with a mean \( \gamma \) correlation coefficient of 0.75 (95% CI: 0.61–0.95) among the 5 observers. The modified lateral pillar and Catterall classifications showed moderate associations (mean \( \gamma \) correlation coefficient 0.55 [95% CI: 0.38–0.66] and 0.64 [95% CI: 0.57–0.72], respectively).

Interpretation — The Catterall classification and the original lateral pillar classification had sufficient interobserver agreement and association to late radiographic outcome to be suitable for clinical use. Adding the borderline B/C group did not increase the interobserver agreement or prognostic value of the original lateral pillar classification.

Several prognostic indices have been proposed in Perthes disease. Major milestones were the introduction of the Catterall and lateral pillar classifications (Catterall 1971, Herring et al. 1992), attempting to predict the final radiographical outcome at an early stage of the disease. Catterall (1971) was the first to emphasize the relationship between the extent of femoral head involvement and final outcome. He defined 4 groups based on the site and extent of femoral head involvement, ranging from less than 25% in group I to a total head involvement in group IV. The classification was developed to be applied in the fragmentation phase. Limitation of the Catterall classification was a difficult and inaccurate initial assessment until the fragmentation phase. Grouping tended to change if the classification was applied too early (Van Dam et al. 1981). Another criticism has been the lack of sufficiently high levels of interobserver agreement (Hardcastle et al. 1980, Christensen et al. 1986, Simmons et al. 1990, Forster et al. 2006).

Herring et al. (1992) introduced a 3-group classification based on the height of the lateral portion of the femoral epiphysis (termed lateral pillar) compared with the unaffected side on AP radiographs. Group A hips showed no involvement of the lateral pillar. Group B hips had lucency and loss of height, but not exceeding 50%. Group C hips exhibited more lucency and > 50% loss of height. Reported limitations of this classification include difficulties to reliably classify hips in the initial stage (Lappin et al. 2002, Kuroda et al. 2009). Another limitation is the difficult use of the classification in bilateral cases since there is a lack of reference height to compare with.

The Herring group (2004a, 2004b) reviewed all the hips in the original study and identified a group of hips with radiographic findings that were more severe than those typical of group B but less severe than those seen in group C. Thus, they introduced a new group termed B/C borderline, transforming...
their 3-group classification into a classification with 4 categories. The good to excellent interobserver results presented by the Herring group for the modified 4-group classification could not be confirmed by recently published results from the UK (Rajan et al. 2013). Thus, the first aim of our study was to assess the interobserver agreement of the modified lateral pillar classification compared with the Catterall and the original lateral pillar classifications.

Besides sufficient interobserver agreement, requirements of a good initial classification include a satisfactory ability to predict long-term outcome. Although the inventors of the modified lateral pillar classification reported good prognostic value (Herring et al. 2004a, 2004b), there seems to exist only 1 later study that has investigated this association (Froberg et al. 2011). Thus, our second aim was to assess the prognostic value of the modified lateral pillar classification and evaluate whether it was a better predictor compared with the Catterall and the original lateral pillar classifications.

Patients and methods
By a systematic search of the radiographic archive of our hospital, we identified 152 children who had been treated for Perthes disease between 1950 and 1984. 139 children had satisfying radiographic follow-up with good visual quality at least 5 years after diagnosis. We selected a random sample of 50 patients using a random-number generator. 5 patients with bilateral Perthes disease were excluded. We used for each patient true anteroposterior (AP) pelvis and frog-leg lateral radiographs at diagnosis, 1-year follow-up (mean interval 14 months) and at 5-year follow-up (mean interval 59 months). Radiographic staging according to Waldenström (1922) was applied. We excluded 3 patients due to advanced radiographic stage (reossification phase). Thus 42 patients (38 boys) with a mean age at diagnosis of 6.5 (3–11) years were included in the present study. At diagnosis, there were 36 patients in initial stage and 6 patients in fragmentation stage. 5 patients had been treated with femoral varus osteotomy and 37 patients had been treated non-operatively.

Observers
5 observers participated in the present study with the following professional background and individual contributions:

Observer SH: specialist in orthopedic surgery, senior pediatric orthopedic fellow.

Observer SS: consultant in orthopedic surgery with a great interest in pediatric orthopedic surgery. He received all radiographs stored on CDs but due to a hardware failure he was only able to retrieve images of 37 patients for the first session. A new set of CDs was sent for the second session and 40 patients could be assessed.

Observer EM: consultant in radiology, with special interest in pediatric orthopedics.

Observer AC: Professor emeritus of pediatric orthopedic surgery. He received all radiographs stored on CDs and found radiographs of 41 patients eligible for this study. No Stulberg classification was applied.

Observer OW: pediatric orthopedic consultant with special interest in Perthes disease.

All observers were familiar with the investigated classifications but nonetheless invited to a consensus-building meeting. All but 1 observer (AC) participated in the meeting before commencing the study. The 4 observers were provided with the original articles and a 20-minute tutorial, outlining the characteristics of each classification.

Radiographic assessment
The radiographs were assessed in 2 separate sessions. In the first session the original lateral pillar classification (Herring et al. 1992) and the Catterall classification (Catterall 1971) were applied, using the radiographs (at diagnosis or 1-year follow-up) that showed the greatest involvement of the femoral head at fragmentation. Radiographic outcome at 5-year follow-up was classified by 4 observers in the 37 non-operatively treated patients based on the shape of the femoral head. We modified the 5-group classification of Stulberg et al. (1981) into a simplified 3-group classification (Wiig et al. 2007), in which group A hips have spherical femoral head, group B have ovoid femoral head, and group C hips have flat femoral head. The second session was at least 1 month later and neither possible marks nor labelling from the first session could be traced on the radiographs. The observers were asked to classify the radiographs at fragmentation according to the modified lateral pillar classification (Herring et al. 2004a).

Interobserver analysis
We included all 42 patients and used an overall kappa statistic assessment of interobserver agreement by calculating the weighted kappa (Cohen 1968) for each pair of the 5 observers, yielding 10 kappa values for the lateral pillar classifications and the Catterall classification. Further, we calculated the weighted kappa for each pair of the 4 observers assessing the modified Stulberg classification, yielding 6 kappa values. Kappa statistics with linear weighting were used, defining the imputed relative distance between ordinal categories as 1 (Lowry 2015). The mean of kappa values for each classification was recorded as the overall kappa value (Light 1971) and they are presented with 95% confidence interval (CI). Possible values for kappa statistics range from –1 to 1, with 1 indicating perfect agreement and 0 indicating random agreement. As suggested by Landis and Koch (1977), we interpreted the weighted kappa values as follows: < 0.20 indicates poor agreement, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good agreement, and > 0.80 indicates excellent agreement.

Association to 5-year radiographic outcome
Only the non-operatively treated 37 were included in this part
of the study. 4 observers classified the radiographic outcome at 5-year follow-up according to the femoral head sphericity using the 3-group modification of the Stulberg classification (round, ovoid, flat femoral head). We termed the category as “true” if there was more than 50% consensus among the observers. If there was ≤50% consensus, the radiographs were reassessed by 2 observers (SH and OW). Loss of height within 2 mm of a concentric circle on AP and frog-leg projection was defined as round and more than 2 mm as ovoid. The associations of the Catterall and the lateral pillar classifications were assessed in a cross-table analysis with “true” Stulberg 3-group classification as outcome variable. Gamma statistics for ordinal variables were used (Goodman and Kruskal 1954, 1959), calculating γ correlation coefficients, which were interpreted as follows: values < 0.24 indicate no association, 0.25–0.49 means weak association, 0.50–0.74 moderate association and values > 0.74 indicate strong association. Statistical analysis was done using SPSS® statistics version 21 (IBM, Armonk, NY, USA).

Results

Interobserver analysis

The kappa analysis (Table 1) revealed that the original lateral pillar classifications had an overall moderate interobserver agreement (mean weighted kappa 0.49, CI: 0.41–0.57). An overall moderate interobserver agreement was also found for the Catterall classification (mean weighted kappa 0.43, CI: 0.26–0.61), with a broader variation for individual kappa values. The modified lateral pillar classification scored lowest with fair overall interobserver agreement (mean weighted kappa 0.40, CI: 0.29–0.51) and individual kappa values ranging from 0.15 to 0.59. The 3-group modification of the Stulberg classification had an overall moderate interobserver agreement with a mean weighted kappa value of 0.50 (CI: 0.28–0.71).

Association to radiographic outcome

There was consensus on the femoral head shape in 32 of 37 patients. In the remaining 5 patients no primary consensus was reached since 2 observers chose “round” and 2 chose “ovoid”. These 5 patients were reassessed by 2 of the observers and the “true” category was agreed upon (Figure). Thus the “true” 3-group Stulberg category was round femoral head in 10 patients, ovoid femoral head in 22 patients and a flat femoral head in 5 patients. The original lateral pillar classification was moderately to strongly associated with 5-year radiographic outcome, with mean (range) γ correlation coefficient 0.75 (0.61–0.95) among the 5 observers (Table 2). The modified lateral pillar had a weak to moderate association with radiographic outcome, with mean γ correlation coefficient 0.55 (0.38–0.66), and the Catterall classification showed moderate association with mean γ correlation coefficient 0.64 (0.57–0.72).

Discussion

Our results revealed moderate interobserver agreement for the Catterall and the original lateral pillar classifications and fair agreement for the modified lateral pillar classification. The original lateral pillar classification applied at fragmentation was strongly associated with the final radiographic outcome assessed by femoral head shape. The introduction of the borderline B/C group did not increase the interobserver agreement or association to late radiographic outcome of the lateral pillar classification system.

Table 1. Interobserver agreement of the Catterall classification, the original and modified lateral pillar classifications and the modified Stulberg classification

<table>
<thead>
<tr>
<th>Classification</th>
<th>Weighted kappa</th>
<th>CI a</th>
<th>Range</th>
<th>Agreement b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catterall</td>
<td>0.43</td>
<td>0.26–0.61</td>
<td>0–0.73</td>
<td>moderate</td>
</tr>
<tr>
<td>Original lateral pillar</td>
<td>0.49</td>
<td>0.41–0.57</td>
<td>0.35–0.72</td>
<td>moderate</td>
</tr>
<tr>
<td>Modified lateral pillar</td>
<td>0.40</td>
<td>0.29–0.51</td>
<td>0.15–0.59</td>
<td>fair</td>
</tr>
<tr>
<td>Modified Stulberg</td>
<td>0.50</td>
<td>0.28–0.71</td>
<td>0.38–0.57</td>
<td>moderate</td>
</tr>
</tbody>
</table>

a CI is 95% confidence interval  
b according to Landis and Koch (1977)

c 3 groups  
d 4 groups

Table 2. Association between the prognostic classifications and the femoral head sphericity at 5-year follow-up assessed by the modified 3-group Stulberg classification. Initials are observer

<table>
<thead>
<tr>
<th>Femoral head sphericity</th>
<th>Correlation coefficient a</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original lateral pillar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.86</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>SH</td>
<td>0.61</td>
<td>0.02</td>
</tr>
<tr>
<td>OEM</td>
<td>0.95</td>
<td>0.001</td>
</tr>
<tr>
<td>SS</td>
<td>0.73</td>
<td>0.006</td>
</tr>
<tr>
<td>OW</td>
<td>0.61</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Catterall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.61</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>SH</td>
<td>0.57</td>
<td>0.006</td>
</tr>
<tr>
<td>EM</td>
<td>0.72</td>
<td>0.001</td>
</tr>
<tr>
<td>SS</td>
<td>0.71</td>
<td>0.002</td>
</tr>
<tr>
<td>OW</td>
<td>0.61</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Modified lateral pillar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>SH</td>
<td>0.66</td>
<td>0.003</td>
</tr>
<tr>
<td>EM</td>
<td>0.38</td>
<td>0.1</td>
</tr>
<tr>
<td>SS</td>
<td>0.63</td>
<td>0.02</td>
</tr>
<tr>
<td>OW</td>
<td>0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean</td>
<td>0.55</td>
<td></td>
</tr>
</tbody>
</table>

a Gamma statistics for ordinal variables.
Before discussing the clinical and scientific implications of these findings, it is important to address the limitations of our study. We did not perform a prior power calculation to identify the minimal sample size required for the interobserver analysis. However, the number of patients was similar to that of previous studies (Tables 3–5). Park et al. (2012) performed a structured approach and determined the need of 36 patients, similar to the number of patients in our interobserver evaluation. 2 observers classified no hips as lateral pillar group A, while the other observers identified only 1 or 2 hips as belonging to group A. It is known from the literature that group A hips are truly underrepresented (< 5%) in comparison with group B and group C in the Perthes population (Herring et al. 2004b, Terjesen et al. 2010). This prevalence problem may cause kappa values to be unrepresentatively low (Byrt et al. 1993). In the evaluation of the prognostic value of the classifications, we included patients who had been treated with non-weightbearing and/or physiotherapy, since none of these methods have been proven to have any effect on the natural history of Perthes disease (Wiig et al. 2008). A limitation with the prognostic evaluation was the relatively small number of patients in this analysis compared with other reports on the natural history (Norlin et al. 1991, Joseph et al. 2003, Terjesen et al. 2010). However, the radiographs in these studies were mainly classified by 1 of the authors alone, which poses uncertainty regarding the reliability of the classification applied. We tried to reduce this uncertainty by multiple readings of the prognostic classifications and by establishing a consensus of the final radiographic outcome. This approach requires a substantial amount of ratings per radiograph, which is only
feasible for a limited number of patients. Thus, the chosen approach posed both a limitation and a strength of our study.

**Interobserver analysis**

In studies on interobserver agreement, it is crucial to specify which statistic was used to compute agreement, i.e. Cohen’s kappa (1960), Fleiss kappa (1971), and intra-class correlation (ICC) (McGraw and Wong 1996) and which variant of the statistics was computed (Siegel and Castellan 1988, McGraw and Wong 1996). The different statistical variants can substantially influence the interpretation of interobserver estimates as shown in the following example: we reassessed a study examining the interobserver agreement of the Catterall 4-group classification (de Billy et al. 2002). The authors presented excellent interobserver agreement using ICC statistics (ICC = 0.94), without clearly stating the variant that was used (missing unit and effect of ICC). We reanalysed the given raw data using linear weighted kappa statistics, yielding 36 pairs of observations and an average kappa value of 0.54 (moderate agreement). Norman and Streiner (2008) showed that ICC (2-way, mixed, single-measures, consistency) is identical to a weighted kappa with quadratic weighting, which tends to result in higher kappa values than when using linear weighting.

Herring et al. (2004a) provided each observer with a 20-minute tutorial before rating the radiographs with the modified lateral pillar classification. Although all participating

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Observers</th>
<th>Mean (range)</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pietrzak et al. (2004)</td>
<td>63</td>
<td>3</td>
<td>0.39 (0.28–0.42)</td>
<td>fair</td>
</tr>
<tr>
<td>Present study</td>
<td>42</td>
<td>5</td>
<td>0.43 (0–0.73)</td>
<td>moderate</td>
</tr>
<tr>
<td>Nathan Sambandam et al. (2006)</td>
<td>44</td>
<td>2</td>
<td>0.44</td>
<td>moderate</td>
</tr>
<tr>
<td>de Billy et al. (2002)</td>
<td>19</td>
<td>9</td>
<td>0.54 (0.36–0.77)</td>
<td>moderate</td>
</tr>
<tr>
<td>Wig et al. (2002)</td>
<td>63–158</td>
<td>3</td>
<td>(0.49–0.62)</td>
<td>moderate to good</td>
</tr>
<tr>
<td>Simmons et al. (1990)</td>
<td>100</td>
<td>4</td>
<td>0.62 (0.50–0.67)</td>
<td>good</td>
</tr>
</tbody>
</table>

*a* weighted kappa

*b* interpretation of kappa values (Landis and Koch 1977): < 0.2 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement;

*c* recalculated with data given in the original article.

---

**Table 4. Interobserver agreement of Herring’s original lateral pillar (3-group) classification in 6 previous studies and the present study**

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Observers</th>
<th>Statistics</th>
<th>Mean (range)</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study</td>
<td>42</td>
<td>5</td>
<td>Weighted kappa</td>
<td>0.49 (0.35–0.72)</td>
<td>moderate</td>
</tr>
<tr>
<td>Podeszwa et al. (2000)</td>
<td>33</td>
<td>5</td>
<td>Cohen’s kappa</td>
<td>0.51 (0.43–0.62)</td>
<td>moderate</td>
</tr>
<tr>
<td>Herring et al. (1992)</td>
<td>32</td>
<td>16</td>
<td>Kappa, unspecified</td>
<td>0.52</td>
<td>moderate</td>
</tr>
<tr>
<td>Akgun et al. (2004)</td>
<td>50</td>
<td>3</td>
<td>Kappa, unspecified</td>
<td>0.53 (0.53–0.54)</td>
<td>moderate</td>
</tr>
<tr>
<td>Wig et al. (2002)</td>
<td>63–158</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.56–0.70</td>
<td>moderate to good</td>
</tr>
<tr>
<td>Pietrzak et al. (2004)</td>
<td>63</td>
<td>3</td>
<td>Weighted kappa</td>
<td>0.65 (0.61–0.70)</td>
<td>good</td>
</tr>
<tr>
<td>Nathan Sambandam et al. (2006)</td>
<td>44</td>
<td>2</td>
<td>Weighted kappa</td>
<td>0.72</td>
<td>good</td>
</tr>
</tbody>
</table>

*a* interpretation of kappa values (Landis and Koch 1977): < 0.2 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement.

---

**Table 5. Interobserver agreement of Herring’s modified lateral pillar classification (4 groups) in 2 previous studies and the present study**

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Observers</th>
<th>Statistics</th>
<th>Mean (range)</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rajan et al. (2013)</td>
<td>35</td>
<td>6</td>
<td>Weighted kappa</td>
<td>0.39 (0.05–0.56)</td>
<td>fair</td>
</tr>
<tr>
<td>Present study</td>
<td>42</td>
<td>5</td>
<td>Weighted kappa</td>
<td>0.40 (0.15–0.59)</td>
<td>fair</td>
</tr>
<tr>
<td>Herring et al. (2004)</td>
<td>20</td>
<td>6</td>
<td>Modified weighted kappa</td>
<td>0.71 (0.49–0.89)</td>
<td>good</td>
</tr>
</tbody>
</table>

*a* interpretation of kappa values (Landis and Koch 1977): < 0.2 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement.
observers in our study were experienced in evaluating radiographs of hips with Perthes disease, they were nevertheless provided with a tutorial. Despite a detailed review of the modified 4 group lateral pillar classification with special attention to the borderline B/C group, we were not able to achieve similar results to those of Herring et al. (2004b). Previous studies have highlighted an increasing reproducibility of the classifications when assessed by experienced observers (Simmons et al. 1990, Podeszwa et al. 2000, Wiig et al. 2002, Kalenderer et al. 2005), but similar interobserver levels could not be reproduced in our study. Many reports assessing the interobserver agreement have been performed at 1 institution only (Nathan Sambandam et al. 2006, Park et al. 2012, Rajan et al. 2013), whilst the present study included 3 different hospitals. It is our belief that the present approach gives a more realistic estimate of interobserver agreement as compared with studies performed at a single institution.

Apart from the complexity of statistical methods and interpretation, studies on interobserver agreement in Perthes disease differ substantially in number of observers and radiographs analyzed. Hence direct comparison of presented results should be undertaken cautiously. Nevertheless, we summarized the results of the most relevant studies assessing the Catterall and lateral pillar classifications using kappa statistics and compared them with our results. We found moderate interobserver agreement in 5 out of 7 studies assessing the Catterall classification (Table 3) and in 5 of 7 studies assessing the original lateral pillar classification (Table 4). The introduction of the borderline B/C group decreased the reproducibility of the lateral pillar classification system in our study (Table 5), which is in accordance with the results of Rajan et al. (2013). They found a fair interobserver agreement (mean kappa = 0.39), similar to our mean kappa value of 0.40. The Herring group (2004b) found in average good interobserver agreement (kappa = 0.71) using a modified weighted kappa analysis (quadratic weighting). The authors attributed only half as much importance in misclassifying the borderline group B/C to its adjacent groups as they attributed misclassifying between groups A, B and C. Since quadratic weighting may increase kappa values artificially if an extra category is introduced within a classification system (Brenner and Kliebsch 1996), this may have led to an unrepresentatively high level of agreement.

Our results showed that the Catterall and the original lateral pillar classifications had moderate interobserver agreement. There is no established common understanding of which degree of interobserver agreement may be necessary or appropriate to define a classification system as satisfactory in clinical practice. Some authors abandoned the use of classifications on the basis of moderate agreement (Christensen et al. 1986) while others redefined them as acceptable (Akgun et al. 2004). However, the fact that the introduction of the borderline B/C group even reduced the reproducibility of the lateral pillar classification raises concerns about the usefulness of this modification.

**Association to radiographic outcome**

We assessed the radiographic outcome on the basis of femoral head sphericity because this is strongly associated with long-term outcome (Mose 1980, Stulberg et al. 1981). Our study revealed significant associations between the 3 classification systems at fragmentation and the femoral head sphericity at 5-year follow up. The original 3-group lateral pillar classification had the strongest association, which is in accordance with previous studies (Herring et al. 1992, Ritterbusch et al. 1993, Farsetti et al. 1995, Ismail and Macnicol 1998, Lappin et al. 2002, Terjesen et al. 2010). All hips that had a flat femoral head shape at 5-year follow up had been classified as group C at the fragmentation phase. The Catterall classification as predictor of radiographic outcome is controversial. It did not correlate well with the final radiographic outcome in some studies (Weinstein 1985, Ismail and Macnicol 1998, Gigante et al. 2002), while others confirmed it as a prognostic factor (Dickens and Menelaus 1978, Meurer et al. 1999, Terjesen et al. 2010). Especially when modified into a 2-group classification, distinguishing between more and less than 50% of femoral head necrosis, the Catterall grouping system was a strong predictor of radiographic outcome (Wiig et al. 2008). Our results confirm a significant association between the Catterall 4-group classification and the femoral head sphericity at 5-year follow up, but the association seemed to be somewhat weaker than that of the original lateral pillar classification.

The Herring group reviewed the hips in their original study and identified a group of hips with radiographic findings more severe than those typical of group B but less severe than those in group C (Herring et al. 2004b). Because of difficulties in defining the borderlines between the groups, a new classification group (borderline B/C) was introduced. The authors found that the modified classification was a strong prognostic factor. Our results and the findings of other authors (Froberg et al. 2011) confirm a significant association between the modified 4-group lateral pillar classification and modified Stulberg as outcome variable. To our knowledge the present study is the first to compare the value of the modified lateral pillar classification with the original classification as predictors. Our findings suggest that the introduction of the new borderline B/C group did not improve the association of the lateral pillar system to the femoral head sphericity at 5-year follow-up.

**Summary**

Our results underline that each of the classifications has its limitations; none is perfect. We think that the original lateral pillar system (3 groups) is the most suitable classification in the early radiographic stages of Perthes disease. It is easier to apply (needs only AP radiographs) and was somewhat better associated with the final radiographic outcome as compared with the Catterall 4-group classification. The introduction of the borderline B/C group increased neither the reproducibility nor the prognostic value of the lateral pillar system, which raises concerns about its usefulness in clinical practice.
SH: contributed to the design of the study, selected appropriate conventional radiographs according to the study requirements, monitored the digitalization process, performed statistical analysis and wrote the manuscript. SS, AC, EM: classified radiographs according to the Catterall, Stulberg and lateral pillar classifications. TT: initiated and contributed to the design of the study. He identified and selected radiographs of Perthes patients from the hospital radiographic archive and participated in the writing process of the manuscript. OW: identified and selected radiographs of Perthes patients from the hospital radiographic archives, contributed to the study design, and classified radiographs according to the Catterall, Stulberg and lateral pillar classifications.

We thank Are Hugo Pripp, Department of Biostatistics, Epidemiology and Health, Oslo University Hospital for valuable support and assistance in refining the statistical analysis and Heidi-Karin Lundlie, secretary at the Department of Radiology, Oslo University Hospital, who helped in digitalizing conventional radiographs and provided observer AC and SS with radiographs on CDs.


