PUBO-FEMORAL DISTANCE MEASUREMENTS IN SCREENING FOR DEVELOPMENTAL DYSPLASIA OF THE HIP

by

Hans-Christen Husum

AALBORG UNIVERSITY
DENMARK

Dissertation submitted 2023

.

Thesis submitted:	20-04-2023
PhD supervisor:	Prof. Ole Rahbek Interdisciplinary Orthopaedics Aalborg University Hospital
Assistant PhD supervisor:	Ass Prof. Rikke Damkjær Maimburg Dept. of Clinical Medicine Aarhus University Hospital
	Prof. Bjarne Møller-Madsen Dept. of Children's Orthopaedics Aarhus University Hospital
	Dr. Michel Bach Hellfritzsch Dept. of Radiology Aarhus University Hospital
	Prof. Søren Kold Interdisciplinary Orthopaedics Aalborg University Hospital
	Ass. Prof. Martin Gottliebsen Dept. of Orthopaedics Aarhus University Hospital
PhD committee:	Prof. Carsten Reides Bjarkam Dept of neurosurgery Aalborg University Hospital
	Prof. Daniel Perry Dept. of Children's Orthopaedics Alder Hey Hospital
	Ass. Prof. Klaus Hindsø Dept. of Orthopaedics Rigshospitalet

PhD Series: Faculty of x, Aalborg University

ISSN: xxxx- xxxx
ISBN: xxx-xx-xxxx-xxx-x

Published by:
Aalborg University Press, Skjernvej 4A, 2nd floor, DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2015

ENGLISH SUMMARY

Screening for developmental dysplasia of the hip (DDH) has been performed in the developed world for several decades. Initially the screening consisted of clinical examinations of hip stability, but with the introduction of hip ultrasound (US) in the 1980s, mass screening of DDH based on hip morphology was made possible.

The advent of US screening also introduced the question of which children should receive a hip US, launching a debate that is still ongoing. Should they be referred for US based on a DDH risk assessment (selective screening) or should all children receive a hip US (universal screening)?

While universal screening has been pursued in some countries, today most screening programmes for DDH apply the selective screening approach citing, among other issues, the increased financial cost of universal screening but also the need for specialised examiners to perform the US screening.

Over the years, evidence indicating an inadequacy of selective screening to reduce late diagnoses or ultimately surgical treatment for DDH, has been reported. Health policy makers are therefore left with the choice of either continuing a possible ineffective screening programme or to pursue universal screening, in a time where health care resources are becoming increasingly scarce.

The present thesis presents a possible middle-ground, by investigating the pubo-femoral distance (PFD) as a more accessible US metric, allowing inexperienced US examiners to perform US screening, thus driving down the cost and personnel requirements of a universal US screening programme.

The PFD method is investigated in three separate studies.

First, the accessibility of the PFD method was assessed in the hands of midwives as novice US users. We designed a training programme with the purpose of documenting the learning curve of novice US users learning the PFD method as well as comparing the measurements to those produced by experienced musculoskeletal radiologists.

Second, the midwives trained in the PFD method were employed in a hybrid selective screening programme for DDH. This screening programme retained the traditional clinical examination and risk factor identification of selective screening while adding universal PFD US as a stand-alone referral criterion. The performance of the traditional selective screening and universal PFD screening was then compared in terms of effectiveness in detecting DDH.

Third, while the PFD method has been documented to correlate to DDH, and consequently correlated to the gold standard US methods used in DDH diagnostics, the degree of correlation had never been investigated. Therefore, using the study population of the hybrid screening programme, we assessed the correlation of the PFD measurement to the gold standard US metrics.

In short, the present thesis aims to provide the argument for the PFD method as a screening tool in locations where universal US screening using the gold standard methods are not feasible.

We found a rapid learning curve, higher efficiency of PFD screening compared to traditional selective screening and a strong correlation to gold standard US metrics. These findings suggest that PFD screening may be a viable alternative to traditional selective screening and an acceptable compromise to universal US screening based on the current gold standard methods.

DANSK RESUME

Screening for hoftedysplasi (DDH) er blevet foretaget i adskillige årtier. I begyndelsen var screeningen baseret på kliniske undersøgelser af hoftens stabilitet, men med introduktionen af ultralydskanning (UL) af hoften i 1980erne, blev det muligt at screene for DDH baseret på hofternes morfologi.

UL screening rejste samtidigt spørgsmålet om hvilke børn, der burde modtage en UL af hoften. Skulle børn henvises til UL baseret på en risikovurdering for DDH (selektiv screening) eller burde alle børn modtage en UL undersøgelse af hoften uanset risiko (universel screening).

Dette spørgsmål udviklede sig til en debat, der stadig er pågående. Universel screening er indført i enkelte lande, men selektiv screening er i dag fortsat mest udbredt, da der henvises til de øgede finansielle omkostninger ved universel screening samt det øgede træk på sundhedsprofessionelle ressourcer.

Med tiden indikerer den tilgængelige evidens dog, at selektiv screening for DDH muligvis er ineffektivt, da enkelte større studier har påpeget at det ikke effektivt reducerer oversete diagnoser eller nedbringer antallet af kirurgiske behandlinger for DDH. Beslutningstagere står derfor over for to valg: at fortsætte den muligt ineffektivt selektive screening eller at gå imod universel screening i en tid hvor ressourcerne i sundhedsvæsnet er knappe.

Denne afhandling søger derfor at præsentere et kompromis ved at undersøge pubo-femoral afstanden (PFD), som et mere tilgængeligt UL mål, der muligvis kan tillade uerfarne UL brugere at foretage UL screening og derved nedbringe omkostningerne ved universel UL screening for DDH, både i forhold til økonomi og påkrævet personel.

PFD-metoden undersøges i tre separate studier i denne afhandling.

Først undersøgtes PFD metodens anvendelighed i hænderne på jordemødre, som vores udvalgte UL novicer. Vi designede et oplæringsprogram med formålet, at dokumentere læringskurven for UL novicer, der oplæres i PFD metoden og samtidigt dokumentere novicernes pålidelighed i PFD opmåling sammenlignet med eksperter i muskuloskeletal radiologi.

Dernæst ansatte vi de oplærte jordemødre i et hybrid selektivt screeningsprogram for DDH. Dette screeningsprogram beholdt de traditionelle henvisningskriterier, klinisk undersøgelse og risikofaktorer for DDH, og benyttede samtidigt universel PFD UL screening som et selvstændigt henvisningskriterie til opfølgende hofte-UL.

Effektiviteten af de traditionelle screeningskriterier, i forhold til at detektere DDH, blev derefter sammenlignet med PFD henvisningskriteriet.

På trods af, at PFD metoden er korreleret til DDH, og derved også de gold standard UL metoder der bruges i DDH diagnostik, er graden af korrelationen aldrig blevet undersøgt. Slutteligt undersøgte vi derfor denne korrelation ved hjælp af populationen fra vores hybrid screeningsprogram.

I sin essens, søger denne afhandling at bidrage med et argument for PFD UL som et screeningsværktøj i samfund, hvor universel UL screening for DDH med gold standard UL metoder ikke er muligt.

Vi fandt at PFD metoden hurtigt kunne læres, at effektiviteten af PFD screening signifikant overstiger traditionel selektiv screening og at PFD har en stærk korrelation til gold standard UL mål. Disse fund indikerer at PFD screening kan være et muligt alternativt til traditionel selektiv screening og et kompromis til universal UL screening baseret på gold standard UL metoder.

ACKNOWLEDGEMENTS

I would like to thank all the people , departments and institutions involved in the truly interdisciplinary studies presented in this thesis. You are far too many to acknowledge in a single paragraph, but without your tireless efforts and continuing input, none of this would have been possible. A special thanks will however go to the following people.

To Bjarne Møller-Madsen for establishing the Danish Paediatric Orthopaedic Research network making this interdisciplinary collaboration possible

To Rikke Maimburg for guiding me through the minefield of interdepartmental communication.

To Michel Hellfritzsch for his patience in teaching me the intricacies of paediatric ultrasound and for scanning my own twin boys.

To Søren Kold for having faith in me, this research project and providing the conditions that allowed me to think freely outside the box while still being solidly grounded by the Interdisciplinary Orthopaedics research group at Aalborg University Hospital.

To my supervisor Ole Rahbek for time and again proving, that I did not make a mistake when walking into his office asking for a research project all those years ago. Ole has always been a supporter of whatever new idea I presented and has been a fantastic supervisor and research partner. But most importantly, I have come to consider him a true friend.

Lastly, to my dear life companion Pernille, who, while I was engulfed in the practicalities of clinical research, continued to support me through my frustrations while also providing a safe-haven for our newborn boys and me. While your ears are surely falling off from hearing about hip dysplasia at this point, know that I could not have done any of this, had it not been for your compassion, support and your loving reminders, that meaning is not found in work and research but rather in the real world where our family is.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF ABBREVIATIONS

DDH: Developmental dysplasia of the hip

CHD: Congenital hip dislocation

US: Ultrasound

PFD: Pubo-femoral distance

BRP: Bone rim percentage

FHC: Femoral head coverage

ICC: Intraclass correlation coefficient

BA: Bland-Altman

ROC: Receiver operating characteristics

NNS: Number needed to screen

AI: Artificial intelligence

CHAPTER 1. INTRODUCTION TO DEVELOPMENTAL DYSPLASIA OF THE HIP

1.1. WHAT IS "DEVELOPMENTAL DYSPLASIA OF THE HIP"?

This deceptively simple question is difficult to answer, as controversies regarding the condition and the relationship between hip dislocation and dysplasia has existed through centuries. The following section is therefore dedicated to clarify what we mean when we say developmental dysplasia of the hip (DDH) and how we arrived at this definition.

DDH today describes growth abnormalities of the acetabulum and/or proximal femur with or without subluxation, luxation or instability of the hip. This definition originated in beginning of the 21st century. Before the connection between hip dislocation and anatomical hip abnormalities was made, the condition was described solely in terms of hip dislocation.

In the year 400-300 BC, the first case description of infant hip dislocation was made by Hippocrates. He accredited it to "*infants undergoing trauma or compression inside the womb* (1) and believed it to be congenital. The first mentioning of the condition in modern times was by Palletta and Dupuytren (Dupuytren, 1847). They coined the term primary or congenital hip dislocation (CHD) in 1847 indicating, by name, that the condition is present at birth. While Dupuytren noted acetabular changes, the connection of hip dislocation to hip dysplasia was first suggested by AM Phelps in 1891. Phelps autopsied a child with hip dislocation who had died from meningitis at age 4½ years and noted: "*The acetabulum is seen to be angular in shape, small, and undeveloped*" and went on to suggest: "*Congenital dislocation of the hip is produced, I believe, by injury at birth; injury in utero, or disease in utero; rhachitis; hereditary influences, and, in exceptional cases, if such there are, by arrest of development of the acetabulum*" (1).

While the proposed causes are non-specific, they are nonetheless reflected in much of what we know of DDH today in terms of injury (the crowding phenomenon of oligohydramnios or breech presentation), the relationship to acetabular malformation and indeed hereditary influences.

In the 20th century, hip dysplasia and hip dislocation were increasingly linked. While there was still ongoing discussion about the name of the condition, the umbrella term CHD was most commonly used. In 1989 Klisic published a brief report stating "*the unsatisfactory nature of the traditional term Congenital Dislocation of the hip has been demonstrated in the course of the last 70 years, by the fact that at least 15 different names have been proposed"* and "*The modern term Developmental Displacement of the Hip is much better. It realistically indicates a dynamic disorder, potentially capable, as the baby develops, of getting better or getting worse […] it embraces all the variants of the disorder (dislocation, subluxation and dysplasia), no matter whether they occur prenatally or postnatally*" (2).

Thus, the ground was laid for Bowen and Kotzias-Neto to define "Developmental dysplasia of the Hip" in their 2006 textbook of the same name. They describe hip dysplasia as referring to "*inadequate development of the hip joint, including the acetabulum, femoral head, or both at the same time.*" (3) which was later refined by Sewell et al as "*a spectrum of pathology, ranging from mild acetabular dysplasia with a stable hip through more severe forms of dysplasia, often associated with neonatal hip instability, to established hip dysplasia with or without later subluxation or dislocation*" (4)

To clarify, while CHD is defined by clinical hip dislocation or instability apparent in the neonatal period regardless of acetabular changes, a diagnosis of DDH, by modern definitions, necessitates morphological changes to the hip joint, which may lead to instability or dislocation, but an unstable hip does not constitute DDH without the aforementioned morphological changes. This distinction is important as will be made clear in the following section.

1.2. HOW COMMON IS DDH?

Prevalence in the clinical screening era (pre 1980s) - CHD

Before the introduction of ultrasound (US) screening in the 1980s, CHD was diagnosed by testing for clinical hip instability (i.e., a positive Barlow or Ortolani manoeuvre or Galeazzi test) or by pelvic radiographs. In the clinical screening, the examiner attempts to dislocate an unstable hip (Barlow), reposition a dislocated hip (Ortolani) or test for limb length discrepancy (Galeazzi) with the child in the supine position and with both hips flexed. The reported prevalence of CHD in this period varies from 0.9% to above 10% even within the same ethnic populations (5–7), which may in part be explained by the low reliability and repeatability of the clinical exam techniques(8,9). However, as CHD and DDH are arguably two distinct conditions, albeit with considerable overlap, the historical prevalence of CHD should not be extrapolated into the era of US screening and DDH.

Prevalence in the ultrasound screening era (post 1980s) - DDH

With the introduction of hip US for DDH in 1983 by Reinhard Graf (10), mass screening of children for dysplastic changes of the hip joint became possible. Additionally, Graf proposed a classification system (table 1), reflecting the different stages of the DDH continuum, based on measurements of the inclination of the acetabular roof (α angle) and the time of examination. The classifications of the hips range from normal (type I hip), to DDH (≥ type IIb hip), to completely dislocated and unmeasurable (type IV hip). This system provided a framework for reporting and comparing DDH diagnoses between studies based on hip US (table 1).

Graf classification	Alpha angle (degrees)	Description
Type I (Mature)	>60	Normal
Type IIa (Immature)	50-60	Immature (<3 months)
Type IIb (Delayed ossification)	50-60	≥ 3 months
Type IIc	43-49	Acetabular deficiency
Type IId (Decentered)	43-49	Everted labrum
Type III (Eccentric)	<43	Everted labrum
Type IV (eccentric)	Unmeasurable	Dislocated

Table 1: Sonographic hip types according to Graf. Adapted from table 6.1 p.41 (11)

The reported prevalence of DDH based on the Graf classification scheme vary across time periods, screening programs and geographical locations, with the latter owing to both cultural factors (i.e. infant swaddling(12)) and genetic predisposition (13). In a recent Austrian study of the largest cohort of universally screened children to date, 10% of screened children had dysplastic changes on hip US and 0.8% had true DDH, defined as a Graf classification ≥ IIb (14).

1.3. TREATMENT FOR DDH

Treatment for DDH depends on the severity of the dysplasia, the instability status of the hip and the age of the child. Active treatment ranges from the application of orthoses or casts to tendon tenotomies, open joint reductions and periacetabular osteotomies (15). For conservative treatment, the optimal treatment window is before the child reaches seven weeks of age (16), as the child ages the success rate of conservative treatment declines and surgical intervention may become necessary. This time-dependency on treatment options makes identification of DDH cases at an early age essential, as it minimizes the treatment intervention and the associated treatment risks for the child (17).

1.4. SCREENING FOR DDH

Despite uncertainties regarding the true prevalence of DDH, it is still considered the most common musculoskeletal disorder in childhood. Together with the time-sensitive nature of treatment, this leads to screening for DDH being widely implemented in the developed world, with the primary aim of reducing the number of late or missed DDH diagnoses.

Screening for DDH can be divided into three general categories:

1. Universal clinical screening: All newborns are screened for DDH using clinical hip examinations which are: the Barlow and Ortolani manoeuvres, The Galeazzi test and a test for limited hip abduction.
2. Selective US screening: newborns at risk for DDH are referred for follow-up hip US based on universal clinical screening and identification of risk factors for DDH. Numerous risk factors have been proposed over the years, commonly used risk factors for DDH screening are(18–20) :
    - Breech position in utero or during birth
    - Family history (1. degree relative)
    - oligohydramnios, clubfoot, macrosomia and multiple births have also been associated to DDH, but a clear risk increase has not been documented.

3. Universal US screening: All newborns are screened for DDH using hip US regardless of risk status or clinical findings.

Selective US screening is predominantly used worldwide and is recommended by large institutions such as the American Academy of Orthopedic Surgeons (18) and the National Health Service in the UK (19). In Denmark, selective US screening was implemented in the mid-2000s.

Universal US screening is currently used in parts of Germany, Austria and Turkey and was recently introduced to Mongolia in the Swiss Mongolian Pediatric Project (21).

While screening for DDH is generally recommended, the evidence surrounding these recommendations is unclear. The uncertainties regarding DDH screening will be presented in the discussion section of this thesis.

1.5. ULTRASOUND DIAGNOSTICS FOR DDH

Over time several hip US methods have been proposed, with the most well-known being the Graf, Harcke and Terjesen methods, which are often used in combination. The following section will focus on these US methods, how they are performed and how they differ from the newer pubo-femoral distance (PFD) measurement.

The Graf ultrasound method

The Graf US method is the current gold standard for DDH diagnostics and screening for children below six months (22). Central to the method and the Graf classification scheme, is the measure of the slope of the bony acetabular roof (α angle). The method also describes a measurement of the cartilaginous roof (β angle), which mainly serves to provide individual differentiation in the case of decentred hips.

According to the methodology developed by Graf, the child is placed in a cradle in the lateral decubitus position while the hip is examined using a linear array probe. To ensure repeatability a "standard plane" must be identified consisting of eight anatomical landmarks of which three are further used to check if the correct sectional plane is produced. The US image is only considered fit for use if all these landmark criteria are fulfilled (figure 1.1).

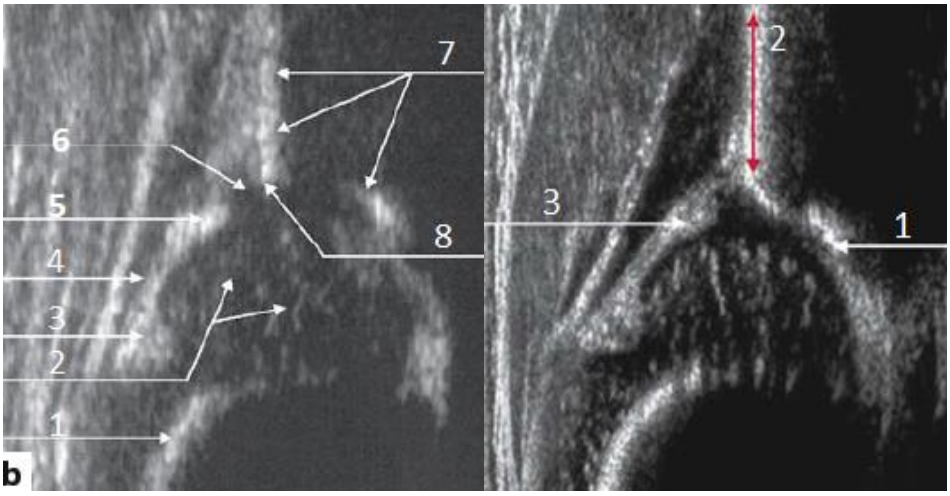


Figure 1.1: (left) ultrasound image of a neonate hip identifying the eight anatomical landmarks for the "standard plane" 1) Chondro-osseous border, 2) femoral head, 3) synovial fold, 4) Joint capsule, 5) Acetabular labrum, 6) cartilaginous acetabular roof 7) bony acetabular roof, 8) bony rim. (right) Ultrasound image of a neonate hip depicting the landmark check for correct plane identification. 1) Lower limb of the iliac bone 2) upper limb of the iliac bone 3) acetabular labrum. Images presented with permission from Hip Sonography, diagnosis and Management of Infant Hip dysplasia by R Graf. (11)

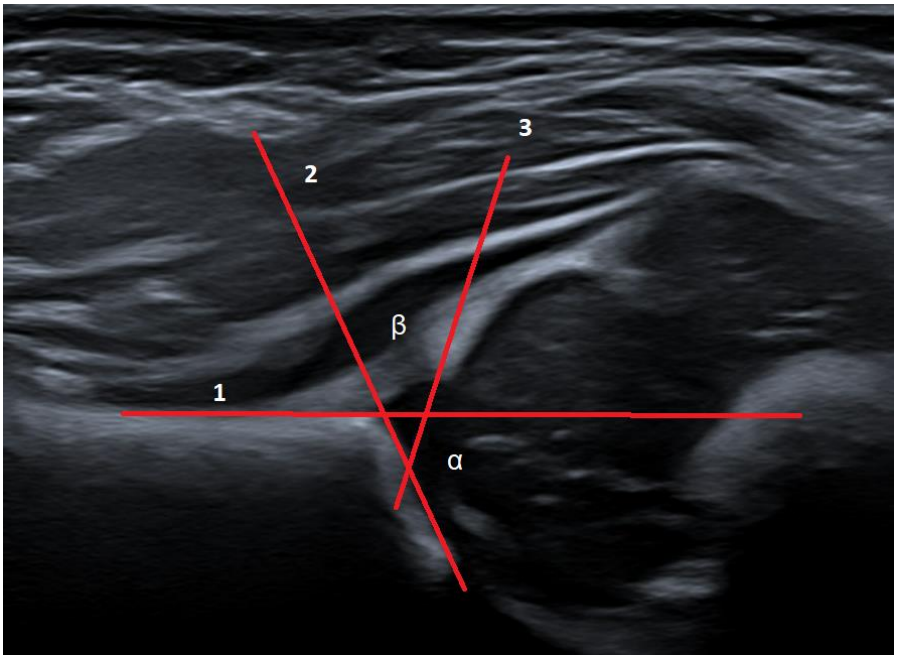


Figure 1.2: Ultrasound image of a neonate hip in the standard plane. 1) baseline, 2) acetabular roof line 3) cartilaginous roof line. α-angle = angle between line 1 and 2, β-angle = angle between line 2 and 3.

Once the standard plane is obtained, the slope of the acetabulum (α angle) is measured as the angle between the baseline, drawn from the uppermost point of the proximal perichondrium tangential to the iliac bone and the bony roof line drawn from the inferior border of the lower limb of the iliac bone to the bony roof (figure 1.2).

The β angle is measured as the angle between the baseline and a line drawn from the "turning point" (i.e., where the lower limb of the acetabulum goes from concavity to convexity) to the centre of the labrum.

Despite its long standing and widespread use, the reproducibility of the Graf method is low (23), which may in part be explained by the complex list of requirements needed in order perform and interpret the Graf US correctly, and the necessary extensive training (24).

This inaccessibility makes an US screening programme, based on the Graf method, difficult to implement on a large scale. However, other, potentially more accessible, US measurements for DDH have been proposed, most recently the PFD method.

The Harcke/Terjesen method

Theodor Harcke and Terje Terjesen both independently described methods for sonographic assessment of hip instability by measuring coverage percentages of the femoral head by the bony acetabular roof. In the original description by Harcke the examination required four views: Two frontal views with the hip joint in neutral and 90° flexion and two transverse views with the hip joint in neutral and 90° flexion (25). The four produced US images in two perpendicular planes are then qualitatively described as normal, subluxated or dislocated. Later, Harcke and colleagues described the optional measurement of the Bony Rim Percentage (BRP) which measures the coverage of the cartilaginous femoral head by the bony acetabular roof in the frontal flexion view (26).

Coinciding with the publication of the Harcke method, Terjesen defined his methodology of assessing the acetabular coverage of the femoral head. In the original description by Terjesen, the method requires static and dynamic examination in the frontal and transverse planes with the hip joint in neutral and slightly flexed position (27). Similar to the Harcke method, a qualitative description of the hip joint is made based on the static and dynamic findings as well as a calculated Femoral Head Coverage (FHC), which represents the coverage of the femoral head by the bony acetabular roof (figure 1.3). While the FHC and BRP both measure the coverage of the femoral head, according to the original descriptions, the BRP requires a frontal 90° degree flexion view while the FHC requires a frontal "slightly flexed" view and the tangential iliac reference line is slightly different for BRP than Terjesen's FHC method. Additionally, the BRP method requires a view of the acetabular fossa, which can be impossible in cases with hip (sub)luxation.

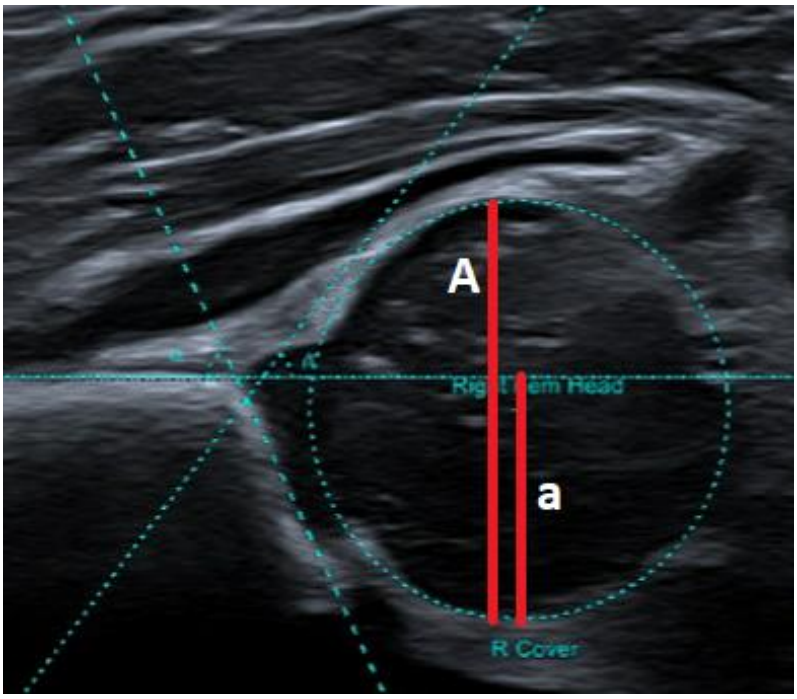


*Figure 1.3: US images of newborn hips depicting the Terjesen FHC measurement (Study III). The FHC is calculated as a/A*100 where a is the distance from the medial edge of the femoral head to the baseline tangential to the iliac bone and A is the distance from the medial to the lateral edge of the femoral head. BRP = Bony Rim Percentage, FHC = Femoral Head Coverage*

Today the term FHC is primarily used and measured according to the method described by Terjesen, although there may be differences in the positioning of the hip joint while performing the measurement. In the present thesis, and study III, FHC will refer to the Terjesen FHC as measured in the frontal 90° flexion view.

The Pubo-Femoral Distance method

The PFD method was first described by Couture et al in 2011(28), although Tréguier et al have reportedly used the measurement since the 1990s (29). According to the methodology described by Tréguier, the child is examined in the supine position with the examined hip flexed in adduction while applying lateralizing stress. The quality criteria for the PFD hip US are: visualization of the epiphyseal centre, the fibrocartilaginous rim, the horizontal upper limb of the iliac bone, the acetabular roof at its greatest depth and the pubic bone (29).

PFD is defined as the minimal measurable distance between the medial margin of the femoral epiphysis and the ossification center of the pubic bone while applying lateralising stress to the hip joint during hip US (figure 1.4). It is therefore similar to the previously proposed Harcke and Terjesen methods, as it quantifies sonographic hip joint instability. However, where the Harcke and Terjesen methods measure the relative coverage of the cartilaginous femoral head by the bony acetabulum, the PFD measures the absolute distance of the femoral head from the pubic bone in millimetres.

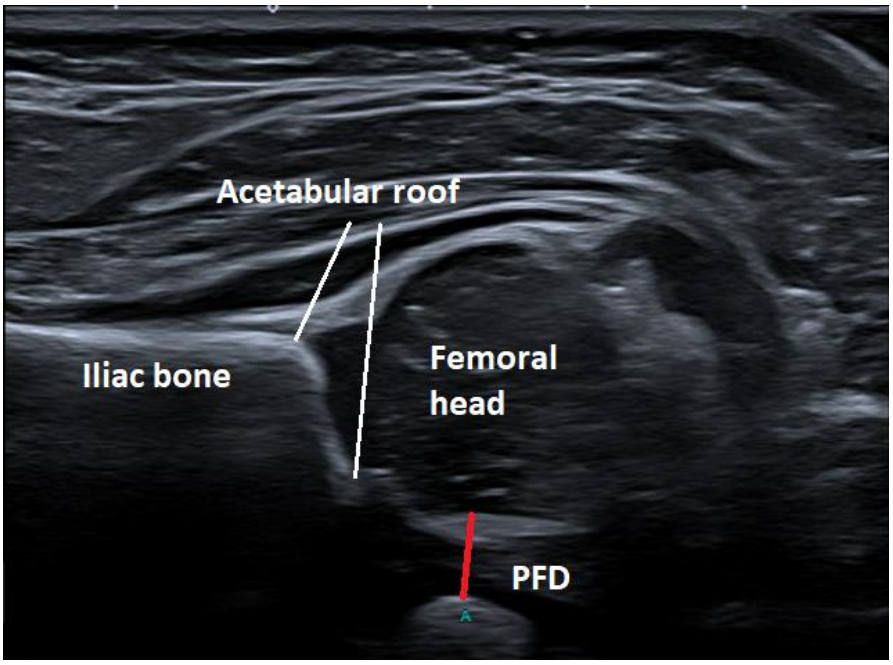


Figure 1.4: US image of a neonate hip depicting the PFD method. The PFD (red) is measured as the minimal distance between the medial margin of the femoral epiphysis and the ossification center of the pubic bone. The quality criteria are: the fibrocartilaginous rim, the horizontal upper limb of the iliac bone, the acetabular roof at its greatest depth and the pubic bone. PFD = pubo-femoral distance.

To date the PFD method is only used systematically in the original authors' region of France, where it reportedly has replaced the Graf method of diagnosing DDH (29). At this point it is necessary to clarify that this thesis and included studies do *not* investigate the PFD as a replacement for the current Graf gold standard, but rather as an alternative to traditional referral criteria in selective screening (i.e. clinical hip examinations and risk factor identification).

1.6. LITERATURE GAPS

Prior to this PhD project, only seven studies had been published about PFD and DDH. From the existing literature the following main points have been reported:

- The PFD has a high sensitivity (100%) and specificity (79.5%) for DDH using a cut-off of 6mm or difference between sides of 1.5mm in children age four to six weeks (29).
- The PFD is reliable between experienced and inexperienced radiologists (Cohen's kappa 0.8, interrater ICC 0.85) (29) (30) .
- PFD increases with age in normal children and is not influenced by gender or side (31).

The available evidence, although limited, may indicate that the PFD method may be the ideal accessible candidate for implementation of mass US screening for DDH. However, more information and confirmation of the above results is necessary. The studies in the present thesis therefore seeks to expand the existing knowledge of the PFD method by investigating three main questions:

1. Can the PFD method be rapidly taught to novice US users?
2. What cut-off value for PFD should be used in early *lateral decubitus* PFD US primary screening for DDH?
3. What is the sensitivity and specificity of the PFD as a referral criterion when compared to traditionally used referral criteria for follow-up Graf hip US
4. How does PFD correlate to existing DDH US metrics?

Regarding question 1

The PFD method has only been documented in the hands of radiologists with varying amounts of training in musculoskeletal US. We wanted to test the limits of the accessibility of the method by training novice US users, with no prior musculoskeletal US experience, in its use. This research question leads to study I of the present thesis.

Regarding question 2

Tréguier et al defined the optimal cut-off value for PFD in their original work. However, the measurements were performed in children at least four weeks old who were examined in the supine position. Further, it is unclear from the 2013 study, how they arrived at their cut-off value and how they defined a true positive outcome to calculate it.

In study II of this thesis, we decided to implement primary PFD screening with as little inconvenience for the child and parents as possible. We therefore mimicked the existing US screening setup of the Graf method by examining the child in the lateral

decubitus position to avoid repositioning of the child during hip US and performed the PFD examination at the same time as the primary clinical screening.

Regarding question 3

No studies have compared PFD screening to clinical hip examination and risk factor identification in early primary screening for DDH. Using the optimised cut-off point for PFD in DDH detection, we aimed to assess and compare the sensitivity and specificity for clinical hip examinations, risk factor identification and PFD as referral criteria for follow-up Graf hip US.

Regarding question 4

The PFD does not directly measure dysplastic changes in the hip joint but rather quantifies sonographic instability. It is however highly sensitive in detecting DDH and must therefore correlate to the Graf scheme of classifying acetabular dysplasia.

In study III we examine the correlation of PFD to Graf's alpha angle and the correlation between PFD and existing validated DDH US metrics that also quantifies sonographic instability.

CHAPTER 2. STUDY AIMS RESUME AND RESULTS

The overall aim of the present thesis was to investigate how the PFD might be implemented in an existing selective DDH screening setting as a referral criterion for follow-up Graf US. The current primary screening for DDH in Denmark is primarily performed by midwives and general practitioners. We chose the midwives as the optimal novice US candidate for training in PFD measurements, as they are a group experienced in connecting with newborns and their parents. Additionally, they may seamlessly integrate PFD US examinations into the logistics of the existing screening setup, at a minimum of inconvenience for both parents and the newborns.

In evaluating the PFD method and the implementation of primary PFD US screening, we focused on the accessibility in terms of learning curves for novice users, effectiveness in detecting DDH with PFD as referral criterion for follow-up hip US, and the correlation of PFD to existing DDH US metrics which we investigated in three separate studies.

The following section will provide a brief resume of each study, the methods used and the results, as well as an overall conclusion of all three studies. This will be followed by an in-depth discussion of the larger methodological challenges faced in the execution of the studies.

The studies are:

- Study I: Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound (published Sep 2022, MDPI Children)
- Study II: Point of care pubo-femoral distance ultrasound outperforms traditional referral criteria in selective screening for hip dysplasia. (Submitted for publication)
- Study III: Pubo-femoral distances correlate to acetabular morphology and sonographic instability in screening for hip dysplasia. (Submitted for publication)

All studies are attached in the appendix section of this thesis.

2.1. STUDY I: PUBO-FEMORAL DISTANCES MEASURED RELIABLY BY MIDWIVES IN HIP DYSPLASIA ULTRASOUND

2.1.1. AIM

To evaluate the reliability and agreement in PFD measurements performed by midwives compared to expert musculoskeletal radiologists and secondly, to investigate the learning curve for the midwives as a group, defined as agreement over cumulative number of hips scanned by the midwife.

2.1.2. RESUME

We recruited midwives for training in the PFD method from the department of Gynaecology and Obstetrics at Aarhus University Hospital (AUH). A training program was constructed based on the theory described in Miller's pyramid (32) as well as utilizing the principles of blended learning in different learning styles and environments. In the training program the midwives were initially presented with a 10 minute online introductory film on the definitions of DDH, how it is diagnosed and treated with an emphasis on the PFD US method. The recruited midwives then participated in a two-hour theoretical seminar which elaborated on the subjects covered in the introductory video. Each midwife and supervising radiologists then participated in a workshop where they measured the PFD on 15 best-practice static US images of newborn hips. Each measurement was repeated at a secondary workshop one week later.

Following this, the midwives participated in four supervised learning sessions, over the course of two weeks. The sessions were integrated into the DDH US screening clinic at AUH, and was supervised by one of three senior musculoskeletal radiologists.

The data gathered was: PFD measured on static US images in workshop #1 and #2 for both midwives and radiologists. PFD measured in the supervised scanning sessions for both midwives and supervising radiologist.

Statistics

Reliability was assessed using intraclass correlation coefficients (ICC) for both inter- and intrarater reliability for the groups of midwives and radiologists. Agreement in the supervised sessions was analysed for each session using absolute measurement differences between PFD measurements, obtained by the midwives and supervising radiologist, as well as inspection of Bland-Altman (BA) plots for any systematic differences in agreement correlated to mean value of PFD measurements.

2.1.3. RESULTS

Eight midwives were recruited for training with a mean seniority of 11 years ranging from 4-27 years. Two midwives had previous experience with foetal US which they had used for one year in daily clinical practice.

Workshop measurements

The reliability of repeated PFD measurements performed on 15 static hip US images was near perfect (ICC>0.89) regardless of group comparators (i.e. intrarater reliability or interrater reliability within the group of midwives/radiologists or between groups (table 2.1).

Table 2.1 – reliability of static PFD measurements performed by midwives and radiologists

	Radiologists (n=3)	**Midwives (n=9)**
Interrater RAD/MW ICC (95% CI) workshop 1	0.99 [0.86;0.99]	
Interrater RAD/MW ICC (95% CI) workshop 2	0.99 [0.92;0.99]	
Interrater ICC (95% CI) within group workshop 1	0.93 [0.84;0.97]	0.89 [0.80;0.95]
Interrater ICC (95% CI) within group workshop 2	0.95 [0.83;0.98]	0.95 [0.90;0.98]
Intrarater ICC (95% CI) group average between workshops	0.98 [0.93;0.99]	0.99 [0.84;0.99]

Table 2.1: inter- and Intrarater reliability of repeated PFD measurements on static US images of newborn hips made by radiologists and midwives.

Supervised live measurements

Included midwives performed a mean of 29 hip USs (range 24-35) over the course of four supervised scanning sessions totalling 237 hip PFD measurements for comparative analysis. Inspection of BA plots did not reveal any correlation between increasing mean PFD values and measurement differences between raters (figure 2.1).

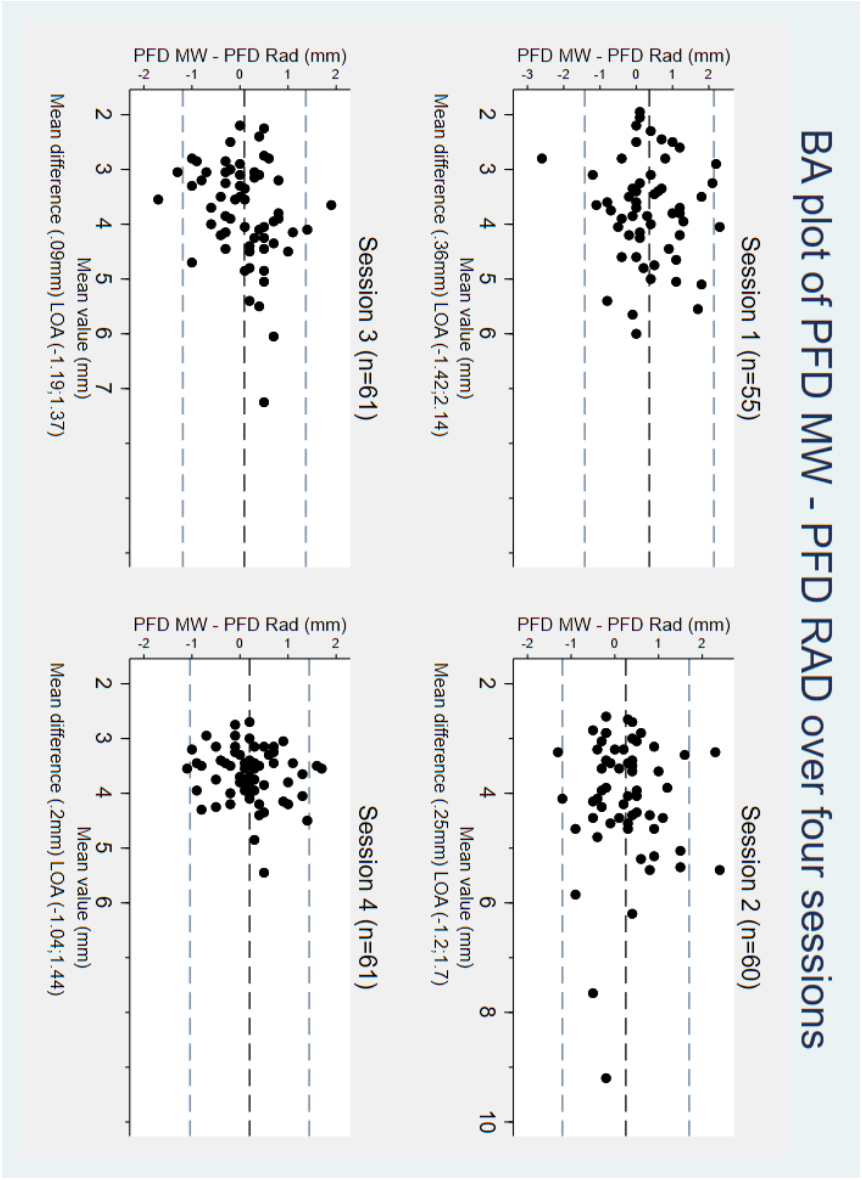


Figure 2.1: Bland Altman plot of PFD measurement differences between supervising radiologist and supervised midwives and mean PFD measurements across four supervised sessions.

Reliability across sessions increased from ICC= 0.59 (95% CI 0.37;0.75) in session 1 to 0.78 (0.66;0.86) in session 3 (table 2.2). We observed a decrease in ICC for session 4 to ICC = 0.42, which did not reflect the observed measurement differences, but may likely be attributable to the low variance of measurement values for session 4, which is a known limitation of the ICC method (33).

Table 2.2	
Session of live scans (n=hips)	**ICC (95 % CI)**
Session 1 (n=55)	0.59 (0.37;0.75)
Session 2 (n=60)	0.81 (0.68;0.88)
Session 3 (n=61)	0.78 (0.66;0.86)
Session 4 (n=61)	0.42 (0.19;0.60)

Table 2.2: Interrater reliability of PFD measurements performed by supervising radiologists and supervised midwives across four supervised sessions.

Initial agreement, as represented by absolute measurement differences between midwife and radiologist, was 0.73mm which decreased by 0.1mm (95% CI 0.02-0.17) for every ten hip US exams the midwife performed. Inspection of a scatter plot of individual- and group mean absolute PFD differences as a function of cumulative number of hip US exams performed by the midwives, supported the abovementioned correlation between agreement and midwife experience. The plot also visualised a decrease in the range of disagreement as the midwives gained experience (figure 2.2).

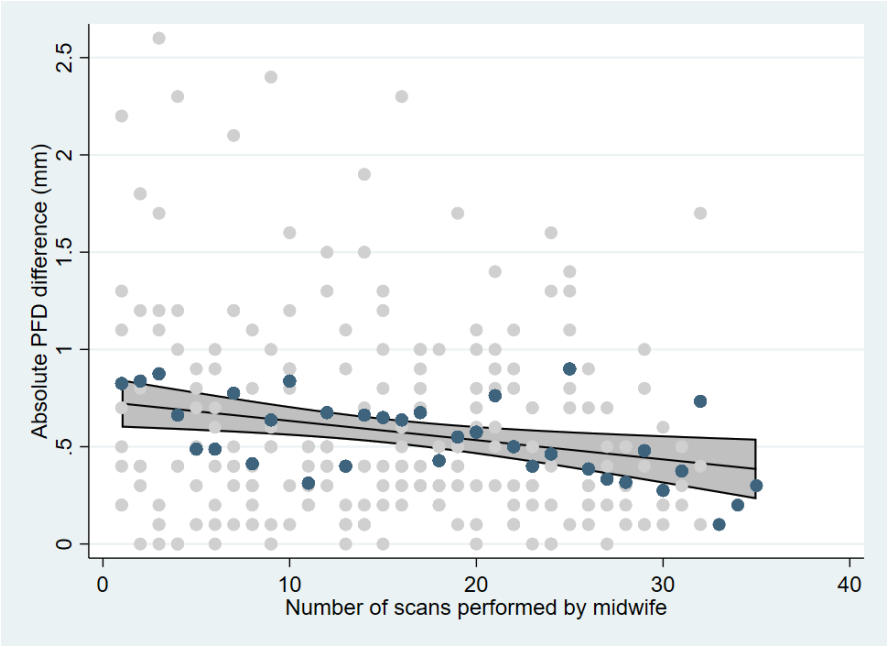


Figure 2.2: Absolute PFD measurement differences between supervising radiologist and supervised midwife as a function of cumulative number of scans performed by individual midwives. Blue dots represent group averages, grey dots represent individual differences. A fitted line with accompanying 95% confidence interval has been added.

2.2. STUDY II: POINT OF CARE PUBO-FEMORAL DISTANCE ULTRASOUND OUTPERFORMS TRADITIONAL REFERRAL CRITERIA IN SELECTIVE SCREENING FOR HIP DYSPLASIA

2.2.1. AIM

To evaluate the performance of PFD measurements as a referral criterion for follow-up hip US in newborns undergoing DDH screening and to compare it to traditional selective referral criteria (clinical hip examination and risk factors). Each referral criterion was compared by sensitivity, specificity and positive predictive values (PPV) in detecting US hip abnormality (≥Graf IIa), the proportion of DDH cases detected and associated referral rates. Additionally, an optimised PFD cut-off value for DDH detection in early PFD US screening was calculated.

2.2.2. RESUME

In the existing screening programme for DDH at AUH, all newborns are screened by a midwife using clinical hip examinations and screening for risk factors for DDH (breech presentation, multiple births, family history of DDH, oligohydramnios, clubfeet) approximately two days after birth in the post-partum clinic. Utilising the midwives trained in PFD US in study I, we offered all parents of newborns, who were screened in the post-partum clinic, a PFD US of their newborn's hips in addition to the clinical- and risk factor screening. The PFD US was performed on the same weekday as the clinical screening, or in the case of newborns clinically screened in the weekends, in the following week.

We included all newborns screened at the post-partum clinic at AUH whose parents opted for a supplementary PFD US, which had to be performed within 14 days after birth. The newborns would be referred for a follow-up hip US if they had a positive clinical examination and/or presence of a risk factor and/or a PFD ≥5.1mm or a difference in PFD between hips ≥ 1.5mm (figure 2.3).

As no other studies have investigated early PFD US examination, the initial cut off value of 5.1 mm for the PFD was established based on a retrospective analysis of PFD measurements of newborns aged 0-14 days from AUH and the available literature (31). It was updated post-hoc with the aim of optimising sensitivities and specificities for PFD as a referral criterion in detecting DDH, and the reported results and figures uses this optimised post-hoc cut-off value.

A consort flow-chart depicting the patient flow for Study II and study III can be seen in figure 2.4. Note that study III uses the initial cut-off for patient selection.

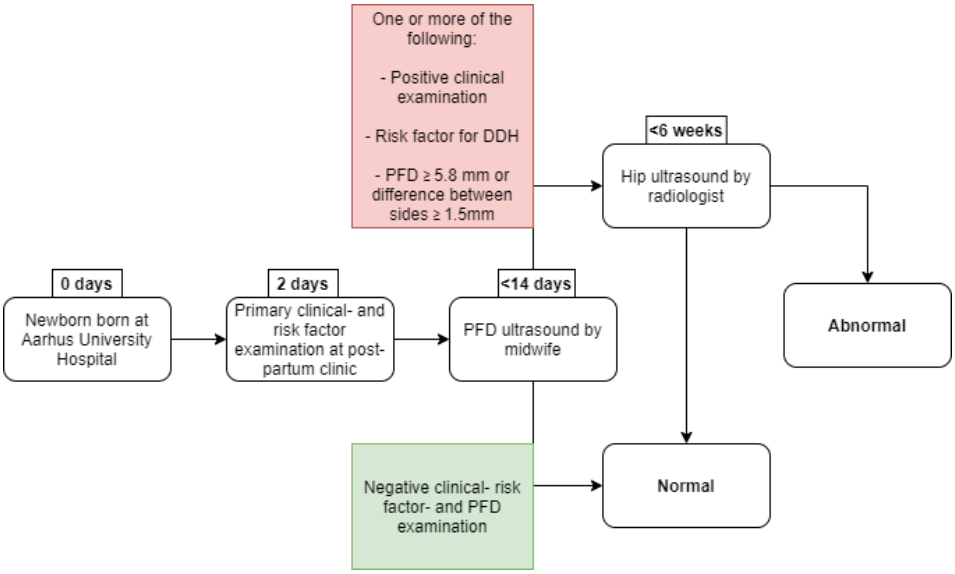


Figure 2.3: flowchart depicting the screening protocol for study II

Statistics

We compared individual referral criteria as well as the combined clinical examination and risk factor identification. Comparison was made using receiver operating characteristics (ROC) curves and AUC values as well as sensitivities, specificities and PPVs with a true positive defined as a Graf hip US classification ≥IIa. Confidence intervals were calculated using bootstrapping (100 samples), while statistical comparison and p-values for sensitivities and specificities were calculated using the Mcnemar test. Observations were investigated for normality using QQ plot inspection and a significance level of 5% was applied.

To optimize the PFD cut-off value in terms of produced sensitivities and specificities empirically optimal cut-off values were calculated using the Youden and Liu indexes (34) (35). The first and senior author decided upon the final value by comparing the empirically optimal cut-off values with a table of cut-off values and their corresponding sensitivities and specificities and chose a value of 5.8mm. A sensitivity analysis of calculated cut-off values using right hips, left hips and all hips was performed to account for bilaterality of observations. As no significant difference in cut-off values was found, independency between hips was assumed.

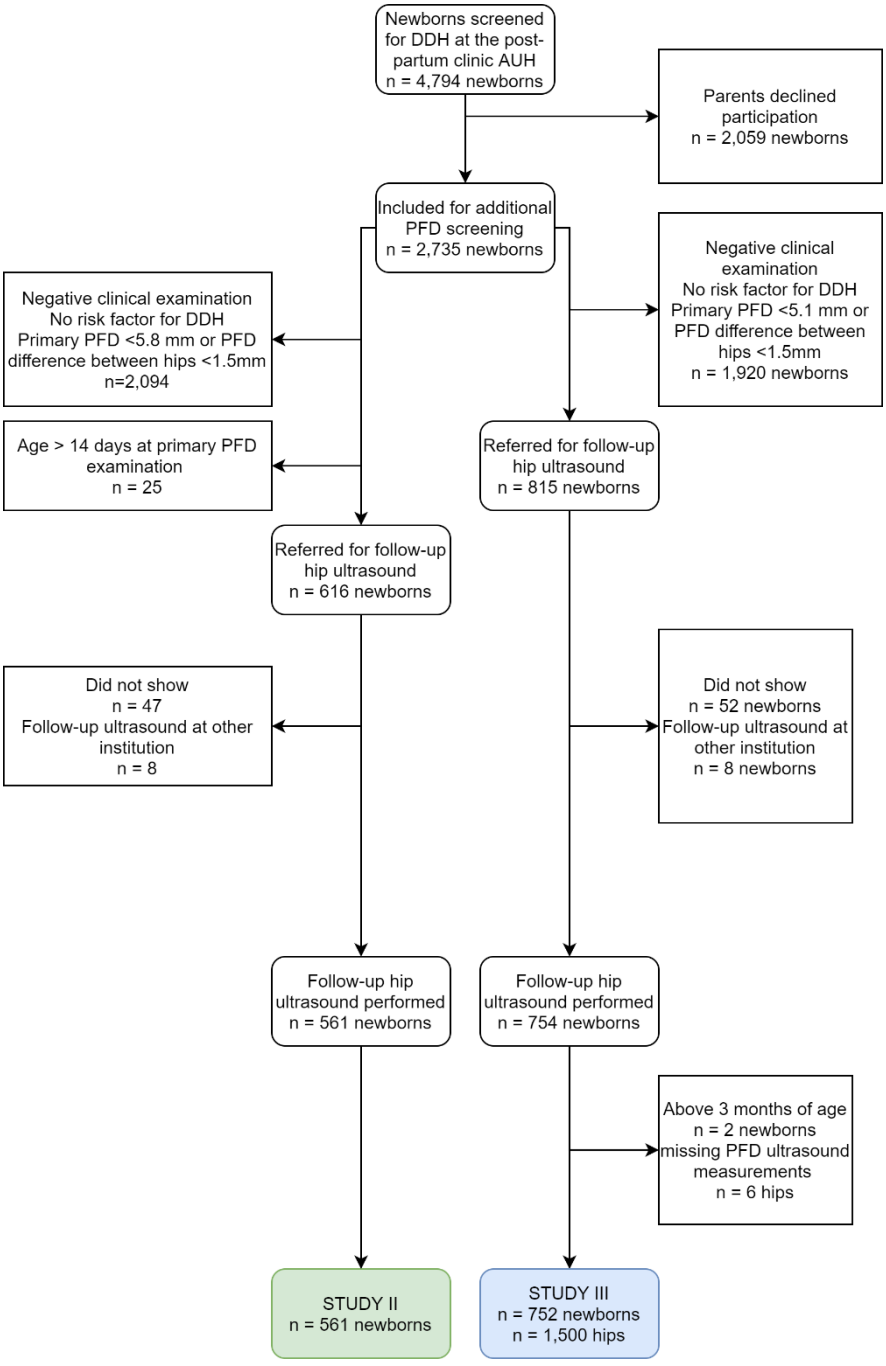


Figure 2.4: combined CONSORT diagram for study II and III using the updated 5.8mm and original 5.1mm cut-off values for the PFD criterion respectively.

RESULTS

Of the 4,794 newborns born at AUH, 2,735 were included for PFD examination after consent was obtained from their parents. Of these, 25 were excluded due to age >14 days at the time of primary PFD examination. 616 newborns were referred for follow-up hip US of which 561 were available for analysis (figure 2.4).

US findings and referral rates stratified by referral criteria (selective or PFD) can be seen in figure 2.5.

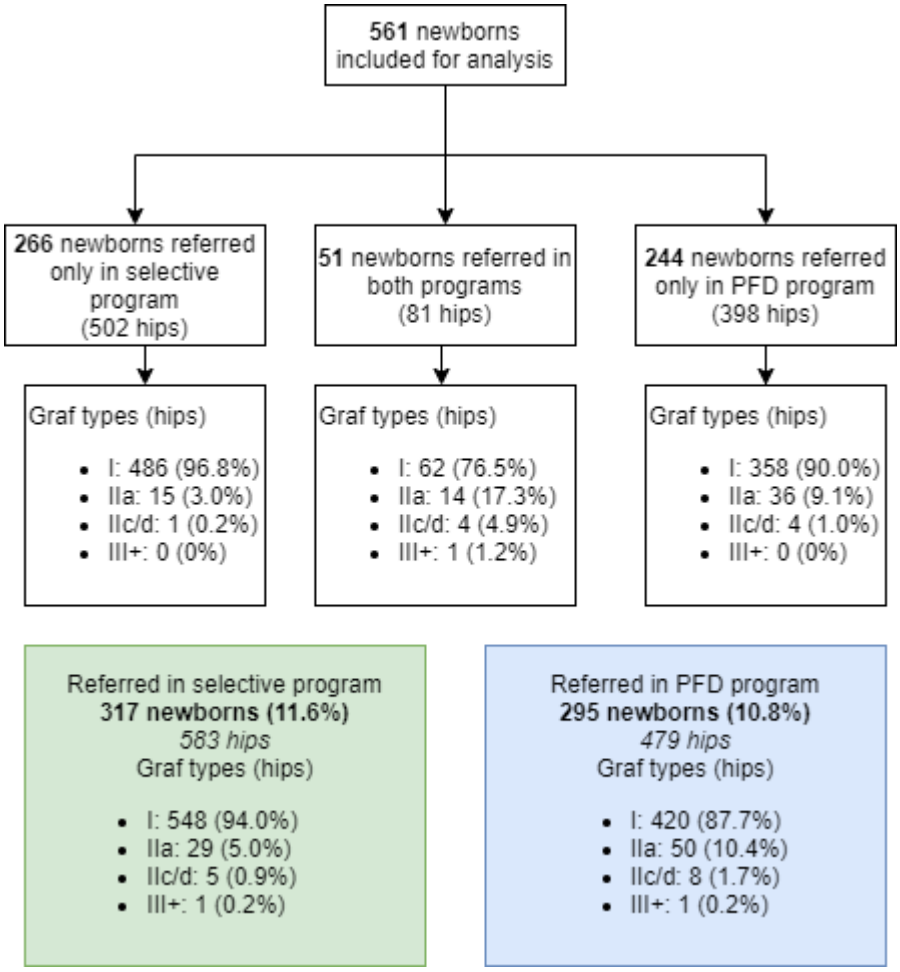


Figure 2.5: Ultrasound findings in the traditional selective screening programme for developmental dysplasia of the hip and the PFD programme. PFD = Pubo-femoral distance.

Calculated sensitivities, specificities, PPV and AUC were significantly higher for the PFD criterion using a cut-off of 5.8 mm when compared to the combined clinical exam and risk factor identification criteria (table 2.3, figure 2.6). The PFD referral criterion detected an additional 21 type IIa hips and three type IIc hips when compared to traditional selective criteria, corresponding to increases of 72% and 60%, when compared to selective programme, while maintaining similar referral rates (11.6% in selective programme vs 10.8% in PFD program). The PFD criterion missed 15 type IIa hips and one type IIc detected by clinical exam and risk factor identification.

Referral criteria	Sensitivity (95% CI)	P-value	Specificity % (95% CI)	P-value	AUC (95% CI)	P-value	PPV % (95% CI)
Clinical examination	17% (10%;24%)	-	94.2% (93%;95%)	-	56% (52%;59%)	-	17% (10%;24%)
Risk factor	28% (19%;39%)	0.15	67.1% (65%;69%)	<0.001	48% (43%;54%)	0.011	5% (3%;7%)
Clinical exam and/or risk factor	41% (29%;50%)	<0.001	62% (59%;64%)	<0.001	51% (45%;56%)	0.049	6% (4%;8%)
PFD (≥5.8mm)	65% (59%;76%)	0.004	78% (76%;80%)	<0.001	72% (68%;77%)	<0.001	15% (12%;19%)

Table 2.3: Performance parameters of referral criteria for follow-up hip US in study II. PFD = Pubo-femoral distance, AUC = Area under the curve, PPV = Positive predictive value.

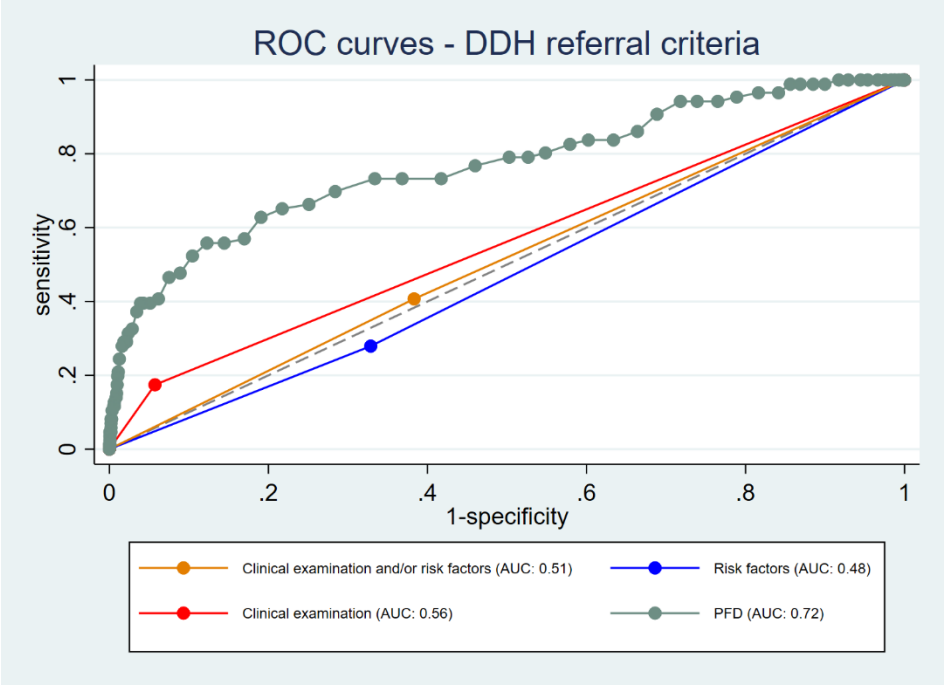


Figure 2.6: Receiver operating characteristics (ROC) curves depicting the sensitivity and 1-specificity of referral criteria for follow-up hip US in study II. The single dots for the traditional referral criteria represent the produced sensitivity and specificity when the referral criteria are positive while the multiple dots for the PFD criterion reflect sensitivities and specificities across different cut-off values for the PFD criterion. PFD = Pubo-femoral distance, AUC = Area under the curve.

2.3. STUDY III: PUBO FEMORAL DISTANCES CORRELATE TO ACETABULAR MORPHOLOGY AND SONOGRAPHIC INSTABILITY IN SCREENING FOR HIP DYSPLASIA.

2.3.1. AIM

To evaluate how the PFD measurement, measured in the lateral decubitus position, correlates to traditionally used US metrics namely Graf's alpha angle and the Harcke/Terjesen FHC in situ and during provocation in hip US.

2.3.2. RESUME

The PFD measurement correlates to the Graf classification scheme (36) and may reportedly decrease the rate of late DDH diagnoses to zero (29). While the connection between PFD and the gold standard Graf method seems evident, to date no studies have quantified the correlation between PFD and the alpha angles which is the foundation of the Graf classifications. Further, no studies have examined how the PFD correlates to US hip stability metrics as measured by the Harcke and Terjesen methods.

We hypothesized that the PFD, although it does not directly measure acetabular morphology, is significantly correlated to Graf's alpha angle as well as the Harcke/Terjesen FHC.

Per the current institutional guidelines, every follow-up hip US for DDH at AUH includes a measurement of alpha/beta angles, FHC in situ and during provocation as well as PFD measurements. During the US examination the child is placed in the lateral examination position, fixed in a cradle. The newborn's hip is flexed in adduction and, using a linear high frequency transducer, the alpha angle, beta angle and FHC is measured according to the methodology described by Graf, Harcke and Terjesen (10,25,27) although FHC was only measured in the frontal plane. The FHC and PFD is then subsequently measured while lateralizing stress is applied to the hip.

We retrospectively collected alpha angles, FHC (in situ and provocation) and PFD measurements from hip US performed on newborns included and referred for follow-up hip US in study II. Exact FHC during provocation was only routinely reported by one of three musculoskeletal radiologists performing hip US, while the others reported the exact FHC value if FHC in situ or during provocation was below 50%.

Included newborns were referred based on a combination of referral criteria which were: primary PFD ≥5.1mm or difference in PFD between sides ≥ 1.5mm, a positive clinical hip examination or presence of a risk factor for DDH. Note that the a-priori cut-off value for PFD of 5.1mm was used in this study, which is reflected in the higher number of referred newborns in study III as compared to study II.

We excluded examinations where the child was above three months of age at the time of follow-up US examination and US examinations where PFD measurements were missing.

Statistics

The correlation of PFD values to alpha angles, FHC and FHC during provocation was analysed using scatter plots and box plots as well as linear regression. To account for bilateral observations a sensitivity analysis using mixed linear models was performed. Mean PFD values were stratified by Graf classification and hip US-displaceability status (displaceable = FHC < 50% in situ and/or during provocation) and compared across strata using Student's t-test. Normality of data was checked using QQ plots while a significance value of 5% was observed.

2.3.3. RESULTS

Of 2,735 screened newborns 815 were referred for follow-up hip US. 53 did not show for follow-up hip US, 8 were referred to another institution, two children were above three months of age at US examination, and six hips were missing one or more US measurement values leaving 752 newborns (1500 hips) for analysis (figure 2.4).

Scatter plot inspection and linear regression revealed a significant negative correlation of PFD to alpha angles, FHC and FHC during provocation ($p<0.001$) (table 2.4 figure 2.7, 2.8). A 1 mm increase in PFD correlated to a -2.1 degree (95% CI -2.3;-1.9) change in alpha angle, a -3.4% (95% CI -3.7;-3.0) change in FHC and a -6.0% (-6.6;-5.5) change in FHC during provocation. PFD was significantly higher for more dysplastic hips, as determined by Graf types and in US-displaceable hips (FHC<50%) ($p<0.001$)(table 2.4 figure 2.7).

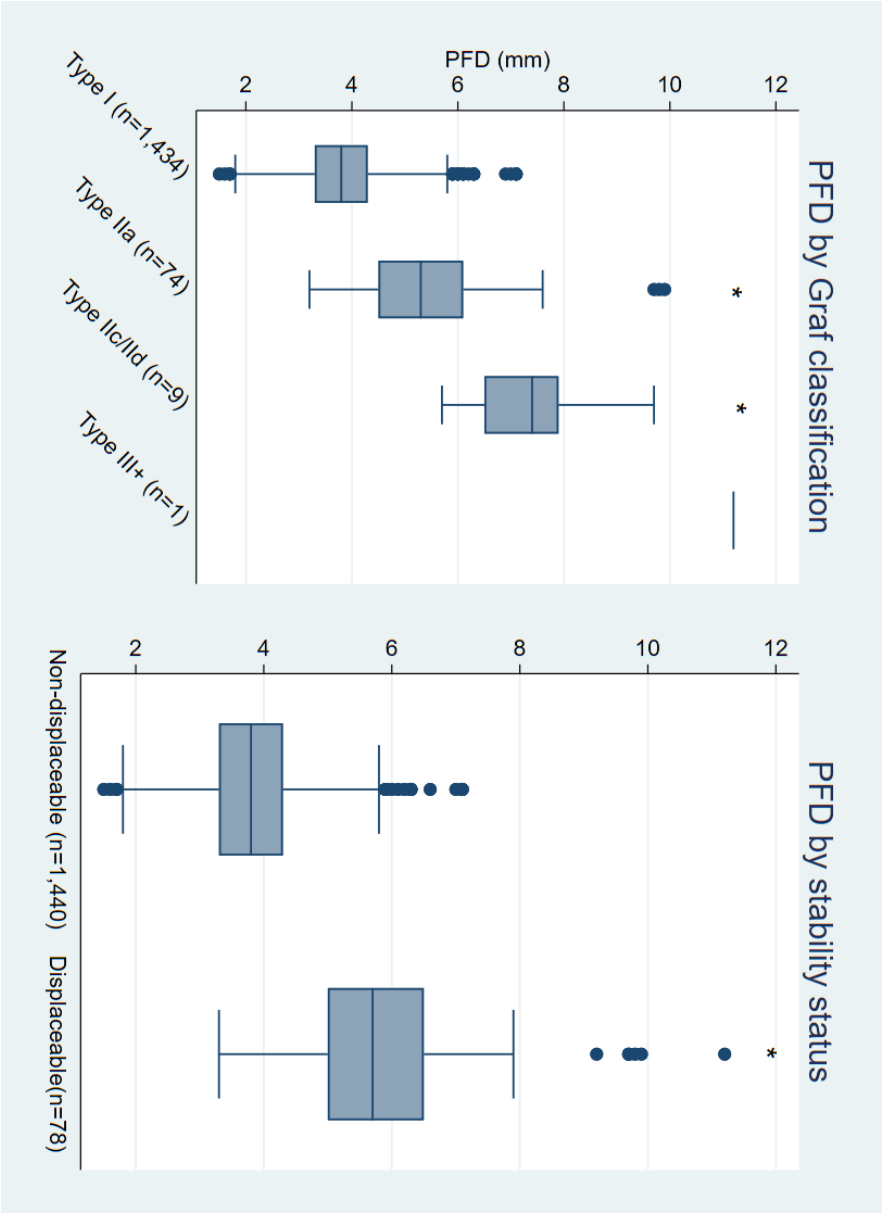


*Figure 2.7: Box plots of PFD values stratified by Graf classification and hip displaceability status. Boxes represent 25%, median and 75% percentiles with whiskers representing upper- and lower adjacent values. Displaceable = FHC or FHC during provocation <50% *=significant result. PFD = Pubo-femoral distance, FHC = Femoral Head Coverage.*

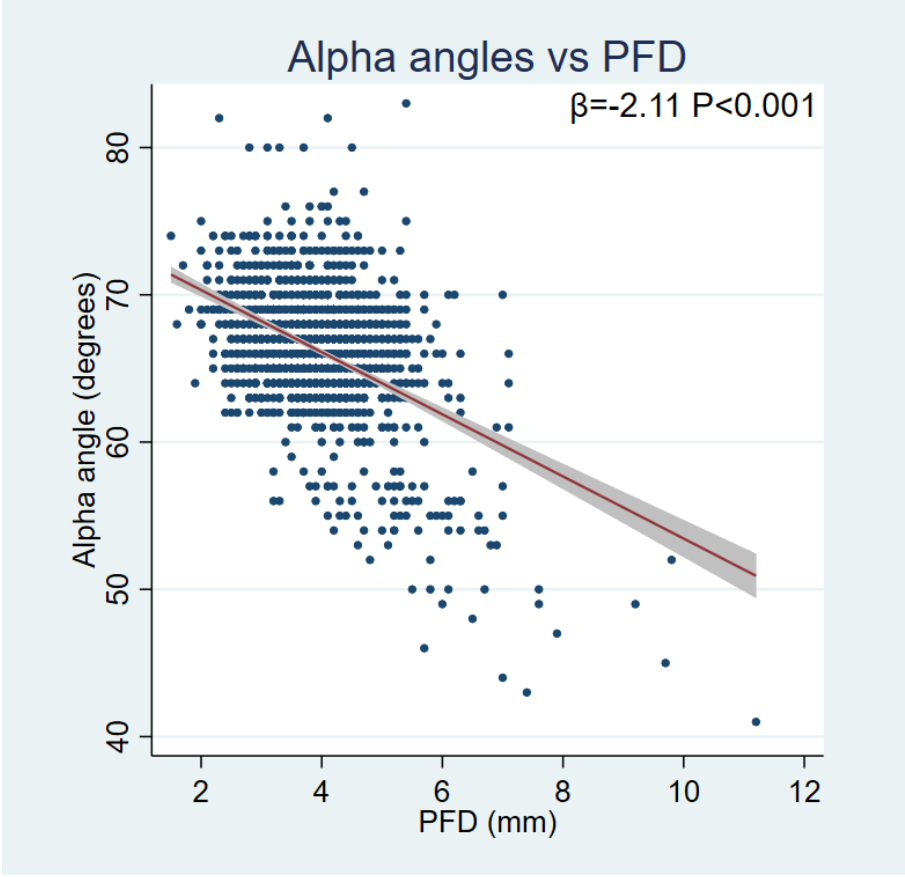


Figure 2.8: Scatter plot of PFD and Alpha angles with fitted regression line, 95% confidence intervals and regression coefficients. PFD = Pubo-Femoral distance.

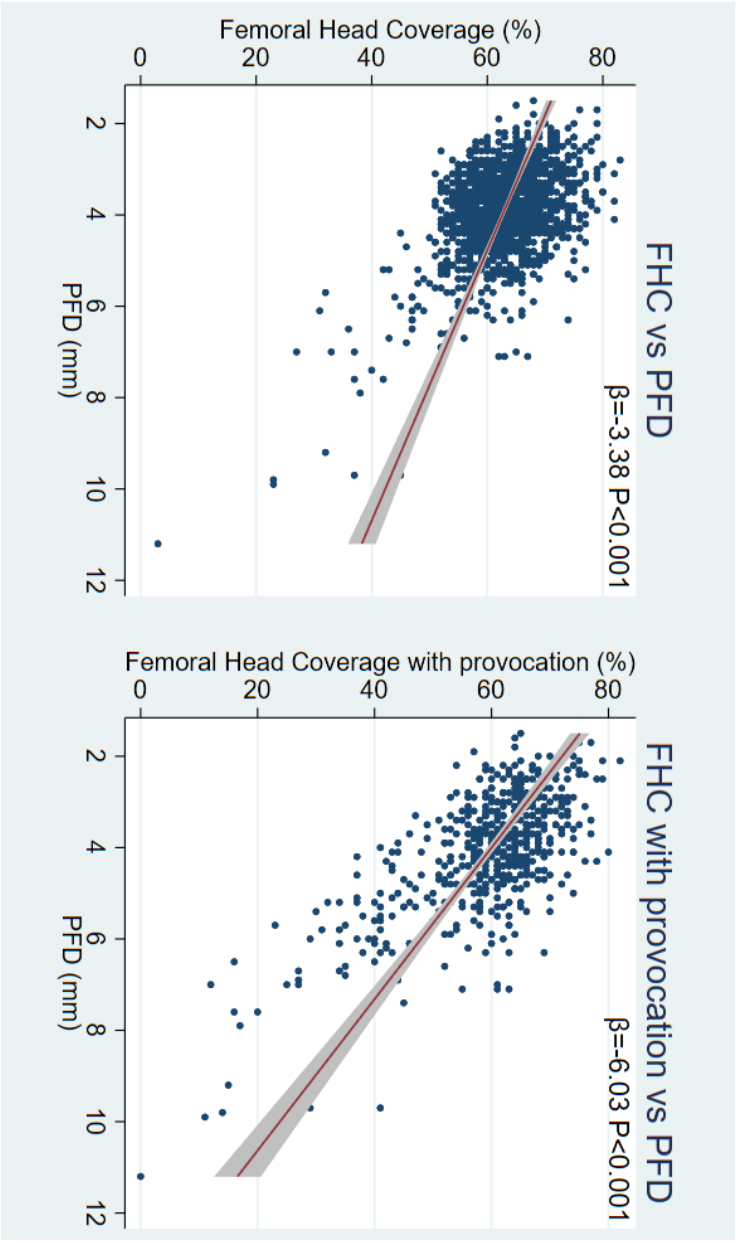


Figure 2.9: Scatter plot of PFD, FHC and FHC during provocation with fitted regression lines, 95% confidence intervals and regression coefficients. PFD = Pubo-femoral distance, FHC = Femoral Head Coverage.

Graf classification (n hips)	PFD (mean 95% CI)	P-value
Type I (n=1,416)	3.8 (3.8;3.9)	
Type IIa (n=74)	5.4 (5.1;5.7)	<0.001*
Type IIc (n=9)	7.4 (6.4;8.5)	<0.001*
Type III+ (n=1)	11.2 (N/A)	N/A
Hip displaceability		
Non-displaceable = FHC>50% (n=1,422)	3.8 (3.8;3.9)	
Displaceable= FHC <50% (n= 78)	5.9 (5.6;6.2)	<0.001*

Table 2.4 distribution of hips according to Graf types and hip displaceability status with accompanying mean PFD values. PFD = Pubo-Femoral distance, FHC = Femoral Head Coverage.

Variable	Intersection	β-coefficient (95% CI)	P-value
Alpha angle	74.5°	-2.1 (-2.3;-1.9)	<0.001*
FHC	76.0%	-3.4 (-3.7;-3.0)	<0.001*
FHC during provocation	84.1%	-6.0 (-6.6;-5.5)	<0.001*

Table 2.5 Results of linear regressions of alpha angles, FHC and FHC during provocation with Pubo-Femoral distance as independent variable. FHC = Femoral Head Coverage.

2.4. OVERALL CONCLUSION

The described studies indicate that:

- The PFD method may be rapidly taught to novice US users producing reliable US measurements. Additionally, the present results suggest that two supervised hip US sessions may be sufficient to produce PFD measurements with low disagreement, when compared to expert examiners.
- PFD $\geq$ 5.8mm as a referral criterion in early primary US screening had higher sensitivity, specificity and PPV in detecting immature/dysplastic hips than clinical hip examinations and risk factor identification combined. Further, it increased detection rates by 60% while maintaining referral rates.
- The PFD measurement is significantly correlated to existing US metrics measuring acetabular morphology and hip displaceability status.

These findings underline the high accessibility, effectiveness and, in part, feasibility of the PFD measurement as a potential replacement for clinical hip examinations and risk factor identification in selective screening for DDH as referral criteria for follow-up Graf hip US.

Specific limitations of the abovementioned conclusions are discussed in the individual supplied manuscripts.

CHAPTER 3. METHODOLOGICAL CONSIDERATIONS

3.1. BILATERAL OBSERVATIONS

One of the central foundations of many statistical methods is the assumption of independency between observations. In health care research, and specifically orthopaedic research, this assumption is often either ignored or violated. In a systematic review from 2006 of 288 studies from seven high impact orthopaedic journals, Bryant et al found that 42% of studies inappropriately used multiple observations (e.g. bilateral joint measurements) from individuals without addressing the potential bias they introduced to their results (Bryant et al. 2006).

In traditionally used statistical analyses, standard deviations, standard errors and associated p-values and confidence intervals rely on the assumption that the random variance in observations originate from the variance between the independent observations of the sample (e.g. patients). However, by including bilateral or within-subject observations, a second source of variance is introduced in the form of possible variance between sides.

Fixed-effects models only take the between-subject variance into account, thus failing to account for the within-subject variance. This is expected to result in an overestimation of the between-subject variance, but an underestimation of the within-subject variance (since this is assumed to be zero in the fixed effect model). It is not possible to know in which direction (wider/slimmer) this error in variance may affect the confidence intervals (37).

As fixed-effects models such as linear regressions, analyses of variance (ANOVA) or generalized linear models only take the between-subject variance into account, we run the risk of underestimating the true variance of our sample (37).

Knowing this, there are several ways to approach the problem of bilateral observations, some of which are mentioned below.

3.1.1. OPTION 1: ASSUME INDEPENDENCE

As Bryant et al found, a prevalent approach is to ignore the risk of bias of bilateral observations by assuming independency between sides. This has the advantage that the methodology may be conducted in a simpler manner and therefore be easier to

understand for non-statisticians. When ignoring independency does not affect the resulting conclusion, simplicity may be an acceptable argument.

In the case of DDH, two joints from the same patient may share certain characteristics because of shared host factors. This could be the increased risk of increased PFD values of both hips because of a shared risk factor such as breech presentation or family history of DDH or, conversely, the absence of any risk factors. Given this correlation, it may therefore not be reasonable to consider observations from two hips from the same patient as independent without performing a sensitivity analysis of how this assumption affects the results and conclusions.

3.1.2. OPTION 2: MIXED MODEL WITH RANDOM EFFECTS

Another option is to use mixed effect models to account for within-subject variance.

The advantage of mixed effects models is the inclusion of more than one source of random variance in data by using random effects. Random effects in mixed effects modelling refer to the within-subject variation that is not accounted for by the fixed effects. For example, in a study of the effect of hip DDH status on PFD values, the fixed effect would be DDH status of the hip whereas the random effect would be any residual variance in data, including the variance introduced by the dependency of within-subject observations (i.e. inclusion of bilateral observations). A mixed effects model would then allow n measurements per subject (with n being either 1 or 2 hips), with one variance estimation for the between-subject observations and another for within-subject observations.

Today many standard analyses are easily done using mixed models in statistical software packages. However, in the case of more complex analyses, some of these tools are no longer easily applied both regarding implementation and time consumption.

In study III, linear mixed effect models were used to calculate the impact of PFD on alpha angles, FHC and FHC during provocation. As the calculated coefficients only differed minimally from the multiple linear regressions, independency between bilateral observations were assumed and the presented results were calculated based on this assumption for simplicity in reporting.

3.1.3. OPTION 3: USING RESAMPLING METHODS.

A third method to account for dependency is to use resampling methods such as bootstrapping. The idea is to draw a new sample from our existing population of same size by randomly picking patients with replacement (38). In this way we preserve the distribution and correlation structure of the sample, but get a slightly different mean

value for each new resample. The quantiles of our resampled means will then give a reliable confidence interval for the estimate in question.

3.1.4. OPTION 4: INCLUDE ONE MEASUREMENT FROM EACH PATIENT

By including only one observation from each patient or collapsing the measurement to a mean for each patient, the problem with bilateral measurements is eliminated. In the case of using mean observations, it is important to recognise if the mean value has a clinical meaning (e.g. changing binary observations to continuous measurements), and to be aware that the interpretation of the result is now about means and not the actual measurement. On the other hand, only one observation is used, a decision must be made about which observation to include. Here an intuitive choice would be to pick one measurement from each patient at random, assuming that there is no systematic difference between measurements depending on side. However, in the case of sides, we may always pick the right side, always the left or simply do both right and left side separately. Whichever is chosen, the researcher must be mindful of the clinical situation as to not bias results by accident and to make sure the clinical interpretation is correct.

An obvious disadvantage of this method is the exclusion of 50% of observations in bilateral data sets, which may therefore not be a tempting choice for the researcher because of the loss of power in analysis. But if the analysis in question can be sufficiently powered with a 50% reduction in observations, this is of little consequence. If sufficient power cannot be retained, that may be a valid argument to exclude this methodology.

An alternative compromise, where statistical power is retained, would be to perform a sensitivity analysis of the impact of assuming independency between bilateral observations on the results. In study II we calculated optimal cut-off values for the PFD in detecting immature and pathological dysplastic hips. The empirical cut-off values were calculated for the right, the left and all hips using the Youden and Liu indexes and compared (table 2.6).

	Only right hips	Only left hips	All hips
PFD cut-off Youden (95% CI)	5.5mm (5.1mm ; 5.9mm)	6.2mm (5.4mm ; 6.9mm)	5.8mm (5.3mm ; 6.3 mm)
PFD cut-off Liu (95% CI)	4.9mm (4.5mm; 5.4mm)	5.8mm (5.4mm ; 6.3mm)	5.7mm (5.3mm ; 6.1mm)

Table 3.1: sensitivity analysis of empirical cut-off values for the PFD criterion calculated according to the methods described by Youden and Liu (34,35) *in study II.*

As the PFD estimates based on one side were within the confidence intervals of the contralateral side, we concluded that the bias introduced by assuming independency between sides was negligible and continued analyses on all hips as independent observations.

3.2. REPRODUCIBILITY MEASUREMENTS – AGREEMENT OR RELIABILITY?

In medical research, reproducibility is often evaluated in terms of agreement or reliability. These terms, while often erroneously used interchangeably (39), are distinct methods of evaluating the reproducibility and repeatability of results. While they both represent a way of looking at the similarity in measurements or scores between examiners, examination methods or grading systems, there are central differences in the way they should be analysed and interpreted.

3.2.1. AGREEMENT

Agreement refers to the degree of similarity or consistency between multiple measurements or evaluations of the same phenomenon. It is generally used to assess the concordance between different methods or observers. In medical research, agreement is used to assess the proximity or precision of measured results between two or more observers or methods in their assessments of the same subject of interest (40,41). In the following example, the agreement analysis from study I, in which we asked midwives and radiologists to measure the PFD of newborns hips, is used.

Agreement was analysed by inspection of Bland-Altman (BA) plots and by quantifying the absolute differences in measurement results for the groups of midwives and radiologists across all examined hips. The Bland Altman plot is constructed by plotting the mean of both observations ($(PFD_{Midwife}+PFD_{Radiologist})/2$) against the difference in observations between raters ($PFD_{Midwife}-PFD_{Radiologist}$). This is done to check the assumption that the measurement differences between raters are not correlated to higher/lower average measurement values i.e. is the measurement difference in PFD between midwives and radiologists affected by increasing average PFD values?

Further, the BA plot inspection would reveal if there were any systematic over- or underestimation of PFD values between raters. In study I, there was no correlation between measurement differences and average measurement values and no systematic over- or underestimation was detected (figure 3.1.

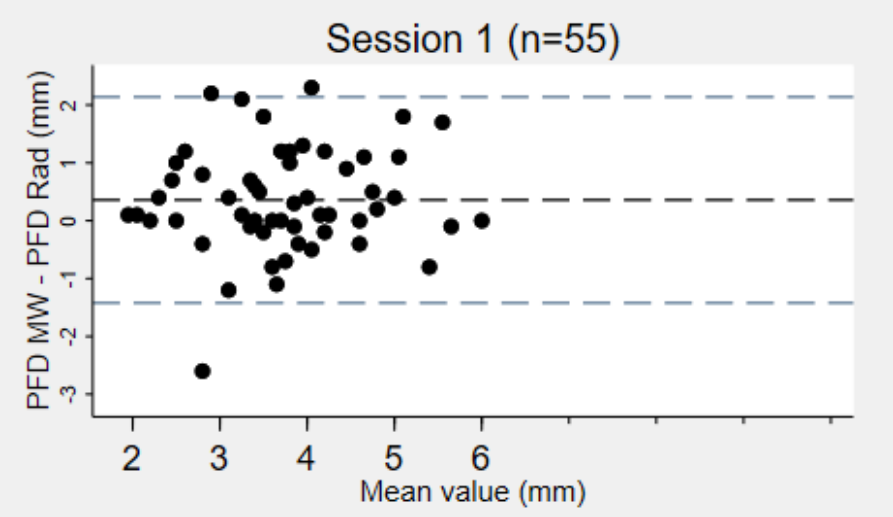


Figure 3.1: Bland-Altman plot of PFD measurement differences between supervising radiologist and supervised midwife and mean PFD measurement values for the first supervised session in Study I.

Having checked this assumption, the agreement can be estimated using standard deviations or average absolute measurement differences. In study I we chose the latter, as a more intuitive way of presenting the disagreement between raters.

The use of absolute measurement differences, as opposed to relative measurement differences is important, when evaluating agreement. In a hypothetical scenario, we

could have observed that there was no systematic trend in the differences between raters, but the midwives both over- and underestimated the PFD values by 100%, when compared to radiologists. Intuitively, this would lead to the conclusion that the agreement is low. But if relative differences were used to evaluate agreement, the mean difference would be close to zero indicating high agreement, when in reality it merely implicates that there is no systematic trend towards over- or underestimation when comparing midwife and radiologist PFD measurements. This is valuable information, and should be reported as a supplement to the BA plot, but it does not give any indications about the agreement between raters.

If absolute differences are used, we get the true measurement error between raters in terms of risk of both over- and underestimation in values relevant for the method in question, as measurement errors are treated equally regardless of the direction of the error.

Therefore, when reporting agreement, absolute measurement differences, combined with BA plot inspection and mean differences should be used (42).

3.2.2. RELIABILITY

Reliability assesses if the subject of study can be distinguished despite measurement errors introduced by the variability between methods/raters (43). Because of the dependency on variability between methods and within observations, reliability measurements are dependent on heterogeneity in the data sample. Commonly used reliability parameters include the Intraclass Correlation Coefficient (ICC) for continuous outcomes, and the Cohen's Kappa value for dichotomous outcomes.

To illustrate the dependency on observational variance, we turn back to the example in study I. Here PFD measurements made by midwives and radiologists during supervised scanning sessions was used, to evaluate the reliability of midwife-performed PFD measurements using the ICC reliability parameter.

ICC calculates reliability using the following basic formula:

$$ICC = \frac{Variance\ in\ observations^2}{Variance\ in\ observations^2 + variance\ between\ examiners^2 + residual\ variance^2}$$

In this example, the variances of interest are the variance in PFD observations and the residual variance introduced by the measurement error between raters. The variance between examiners concerns systematic differences in interactions between examiners and the subject of study (i.e. different settings, locations).

If no systematic interaction between examiners and examined hips, the above formula can be rephrased to:

$$ICC = \frac{Variability\ in\ PFD\ between\ newborn\ hips}{Variability\ in\ PFD\ between\ newborn\ hips + Measurement\ error}$$

From the formula, it follows that the ICC or reliability reflect the variability/variance/heterogeneity of the sample. If the measurement error is high relative to the variability of the observations, we get a low ICC. It therefore follows that, if the variability between observations is low, compared to the measurement error, we get a high ICC (33).

This may seem like trivial algebra, but it is important to recognise the impact of variability in observations when estimating reliability, as this can affect the calculated outcome which was the case in study I.

In study I the reliability of midwives vs radiologists was calculated in terms of their performance of PFD measurements for each session by ICCs using the above formulae. While the agreement increased by each session, the ICC dropped from 0.78 to 0.42 from session three to session four. This drop in reliability was not reflected by the regression analysis of agreement (figure 2.2) or the BA plot inspection (figure 2.1). By inspecting the BA plot for session four, a decrease in variability of PFD measurements was detected when compared to the previous three sessions. As there was no observed drop in agreement with increasing sessions, the decreased ICC in session four was attributed to lower variability between observations, which is a known limitation of the methodology (33).

CHAPTER 4. DISCUSSION

The PFD method is at the centre of the present thesis. To provide the reader with the context and justification for pursuing PFD as a screening tool for DDH, the following section will focus on challenges facing current screening practices for DDH and how the PFD as a referral criterion in selective US screening might improve screening for DDH.

4.1. IS THERE A NEED FOR DDH SCREENING?

This question may seem rather provocative, since DDH screening has been a part of routine health care in developed countries for the past 40 years. However, the effects of DDH screening, and selective DDH screening in particular, are unclear. The central issue is a lack of conclusive evidence in terms of the effect of screening, who to treat when detected and what long term consequences untreated DDH has for the individual (44–46).

Despite the lack of evidence, millions of children worldwide still undergo screening for DDH every year. While the screening itself poses no risk for the child, the risk of false positives may lead to overdiagnosis, overtreatment or cause unnecessary worries for the families involved. In a report published by The United States Preventive Services Task Force in 2006, citing the lack of evidence, they were "*unable to assess the balance of benefits and harms of screening*" and thus could not recommend screening for DDH (47).

In a similar remark, the UK national screening committee took a clear stand on DDH screening in 2004, when they stated:

"*If proposed now as a new programme, DDH screening would probably not be accepted. However, it is so ingrained in the clinical practice of so many people that it would be almost impossible to stop it unless overwhelming evidence of ineffectiveness could be obtained.*" (48)

Even if effectiveness could be obtained and demonstrated, there is still limited agreement on which patients to treat once detected, as the US preventative task force noted in their 2006 recommendation statement:

"[…] *60% to 80% of the hips of newborns identified as abnormal or as suspicious for DDH by physical examination and >90% of those identified by ultrasound in the newborn period resolve spontaneously, requiring no intervention.*" (47)

The question is therefore: how many children with DDH, needing treatment, are born each year?

To answer this, the hips of interest need to be defined which we will define as hips that would eventually require an intervention to prevent residual acetabular dysplasia as opposed to those that will resolve on their own.

There is no universal consensus nationally or internationally on which DDH cases to treat, barring dislocated hips (49,50). The disagreement reflects the lack of conclusive evidence regarding indications for treatment of DDH and when to initiate treatment. In the treatment principles proposed by Graf, treatment should be reserved for decentred hips (Graf types D, III and IV), unstable hips (IIc unstable) and stable dysplastic hips (IIc stable, IIb, IIa-) (11), while immature Graf type IIa hips should be monitored as 5% of will not mature sufficiently within the first three months of life (51).

Two RCTs have investigated these principles of treatment. In 2010, Rosendahl et al randomised 128 newborns with stable Graf type IIc hips to active surveillance or immediate treatment and reported that 45% of the active surveillance group resolved spontaneously without treatment(52). In 2020 Pollet et al randomised 137 newborns with stable type IIb/IIc hips to active surveillance or immediate harness treatment and found that 80% of hips in the surveillance group matured spontaneously (53). However, as only two newborns with type IIc hips were included in the surveillance group, these results should only be considered valid for stable type IIb hips.

What then is the number needed to screen to detect hips eventually needing treatment based on these findings?

As mentioned in the background, the most accurate assessment of the prevalence of hip dysplasia comes from a cohort of universally screened children in Austria(14). In this study, the distribution of hips according to the Graf classification was as follows:

Graf classification	N patients (%)
I	25,093 (90.2)
IIa+	2,467 (8.9)
IIa-	30(0.1)
IIb	9(<0.1)
IIc	182(0.7)
III	19(0.1)
IV	8 (<0.1)
Total	27,808 (100)

Table 4.1: distribution of baseline classification of patients in a universal US screening program for DDH, adapted from Biederman et al 2018 (14)

If an assumption is made, that these prevalence estimates represent unilateral stable cases of DDH, the potential number of cases needing treatment can be estimated by combining the prevalence estimates with the results from the Pollet and Rosendahl RCTs.

In Denmark, 60.000 children are born each year translating to 5.340 type IIa hips, 393 type IIc hips and 58 ≥type III hips. Of the immature type IIa hips, 267 will not undergo spontaneous resolution before three months of age, thus converting them to type ≥IIb hips of which 53 (20% without spontaneous resolution) require treatment (52). Of the 393 type IIc hips, 177 hips will require treatment (53). As there are no studies comparing treatment of type III hips to observation, the need for treatment for these hips is assumed.

Thus 288 DDH hips are in need of treatment each year in Denmark according to the findings by Pollet and Rosendahl.

If, on the other hand, the treatment principles by Graf were followed, all of the type ≥IIc hips would require treatment along with 5% of the monitored type IIa hips which did not resolve themselves, totalling 718 DDH hips needing treatment each year in Denmark.

In a universally screened population, the number needed to screen (NNS), to achieve treatment of the above DDH cases, can then be calculated as the number of patients needed to be screened to prevent one event, in this case DDH requiring treatment. According to the Graf treatment principles, NNS = 60,0000 newborns/718 DDH cases needing treatment = 84, while available evidence suggests NNS = 60,000 newborns/288 DDH cases needing treatment = 208 children. To put these numbers in perspective, the NNS in the Danish screening programme for congenital hearing loss is 400 (54). In terms of NNS, screening for DDH therefore seems reasonable, even when applying the liberal treatment principles proposed by Graf.

However, as it is not yet possible to determine which hips will resolve spontaneously, 9.8% of primarily screened children, assuming perfect sensitivity and specificity of the screening program, will need serial follow-up hip US or pelvic radiographs to distinguish between hips that spontaneously resolve and those that require intervention. It is therefore vital to consider the effectiveness of the primary screening, in terms of sensitivity and specificity, as to not add false positives to the significant proportion of children who need serial follow-up.

4.2. THE EFFECTIVENESS OF SELECTIVE- AND UNIVERSAL US SCREENING FOR DDH

Uncertainties exist regarding the value of early detection of DDH, as a significant proportion of dysplastic hips will resolve on their own without intervention, making the value of detecting these hips questionable(47). It is however widely accepted, that the prevention of late detection of DDH should be the primary goal of any screening program, as late diagnosis increases the risk of surgical interventions and sequelae for the patient (17). This is also reflected by the choice of late diagnosis as the primary outcome in available meta-analyses.

To date, four meta-analyses have been published concerning selective vs universal US screening for DDH, all of which use late diagnosis as a comparative outcome (table 4.2).

Author (year)	Study type	Studies included	Outcome	Result
Shorter et al (2013) (44)	Meta-analysis	5	Multiple, including late diagnosis of DDH	Inconclusive
Jung et al (2020) (46)	Meta-analysis	5	Late diagnosis (6 weeks- 6 months) of DDH	Significant decrease with universal US screening
Kuitunen et al (2022)	Meta-analysis	76	Multiple, including late diagnosis (>12 weeks) of DDH	Nonsignificant effect
Cheok et al (2023)	Meta-analysis	31	Late detection (>3 months) of DDH	Trend towards decrease with universal US screening

Table 4.2: Meta-analyses of selective vs universal US screening for DDH from a literature search of Pubmed and EMBASE using the query: ((Developmental dysplasia of the hip[mesh]) AND (Mass Screening[mesh])) AND (Meta analysis) March 2023.

The systematic reviews and meta-analyses have been complicated by the large heterogeneity of available studies in terms of variations in the application or methodology of selective or universal US screening (e.g. timing of US, clinical examinations used, technological limitations in US), lack of power in analysis, inappropriate study design, geographical variations in the prevalence of DDH, definitions of DDH and comparators used (44). Although there is an abundance of available literature, the lack of sufficiently powered comparative studies leads to a limited number of studies being included for analysis in four out of five of the available reviews. The inconsistent results of the studies mentioned above, may therefore reflect the strictness employed by the authors, when deciding on inclusion criteria and study selection, rather than the true effect of universal- versus selective DDH screening.

The effect of the strictness of inclusion criteria on the effects found is also reflected in the authors' conclusions:

"There is insufficient evidence to give clear recommendations for practice. There is inconsistent evidence that universal ultrasound results in a significant increase in treatment compared to the use of targeted ultrasound or clinical examination alone. Neither of the ultrasound strategies have been demonstrated to improve clinical outcomes including late diagnosed DDH and surgery. The studies are substantially underpowered to detect significant differences in the uncommon event of late detected DDH or surgery." Shorter et al 2013, Cochrane review of five RCTs (44)

As opposed to the more direct conclusion:

"*Reported rates of early-detected DDH and initial nonoperative treatments are higher in settings with universal ultrasonographic screening compared with clinical screening and selective ultrasonographic screening programs. However, the incidences of late detected DDH and surgical treatment rates were not significantly different among different screening strategies.*" Kuitunen et al 2022, systematic review and meta-analysis of 76 studies from 1987 to 2020 (55).

The 2022 meta-analysis by Kuitunen et al did not find any significant differences in late detection of DDH or surgical treatment rates between screening strategies. However, the five conducted meta-analyses and forest plots all have high levels of heterogeneity (I^2 = 85%-100%), indicating large variability in the results of included studies (56). While the authors performed a random effects meta-analysis, to account for unexplainable heterogeneity, they did not perform a thorough investigation of known causes of heterogeneity or sensitivity- or subgroup analysis of their included studies as recommended by the Cochrane Collaboration (57). Care should be exercised when drawing conclusions from extremely heterogenous meta-analyses as the results may be biased in an unknown direction by the inconsistency in the extracted data (56). While the nonsignificant result may likely be attributed to heterogeneity, results from the included studies trended towards a reduction in late DDH diagnosis when universal US screening was employed. In the 21 included studies of selective US screening, late DDH diagnosis incidences varied from 0 per 1,000 children (58–60) to 13 per 1,000 children (61). Conversely, in 11 of the 12 studies reporting on rates of late DDH diagnoses in universal DDH screening, incidence rates varied from 0 per 1,000 children to 1.3 per 1,000 children.

In Austria, universal US screening for DDH has been national health policy for over 30 years (62) representing the longest experience with universal US screening in the world. In the results from the Austrian screening programme published in 2018, none of the 28,092 universally screened children, from one institution, were referred later for treatment in the five year study period (63).

Conversely, 26 years of selective US screening in the UK has failed to reduce the rate of late diagnoses of DDH (1 year or older) (64), and 85% of American patients with symptomatic adult acetabular dysplasia did not have clinical signs or risk factors for DDH at birth (65). Even in the case of the expert clinical screening in Norway, where no effect of universal US was found in RCTs, only eight percent of patients receiving a hip arthroplasty, because of hip dysplasia, had a positive clinical examination at birth (66). It may be that the hips detected through selective US screening are not the ones at risk of persistent dysplasia.

The findings from the American and Norwegian arthroplasty patients concern adult acetabular dysplasia, which until recently had not been definitively linked to paediatric acetabular dysplasia. Laborie et al established this link in their follow-up study of one of the Norwegian RCTs (67). In the follow-up study they performed pelvic x-rays on 2,370 adult patients aged 18-19, who were included in the previous RCT, and found that alpha angles and US hip instability measured in infancy, were significant predictors for acetabular dysplasia in adulthood (68). Additionally, clinical stability assessment and family history of DDH did not predict adult dysplasia, further problematizing selective US screening strategies, which rely heavily on both. These findings provide additional significance to the findings describing the inadequacy of selective US screening in preventing adult dysplasia.

Higher early DDH detection rates in screening leads to increases in potentially unnecessary treatments (69). But as it is presently not possible to discriminate between spontaneous self-resolvers and those at risk of persisting dysplasia, it may be necessary to perform early detection and follow-up of all DDH cases to lower late detection rates. Universal Graf US screening programmes increases rates of early detection and referrals (69), but as the results from study II suggest, early detection rates can also be significantly increased by adding accessible US methods into the existing selective US screening setup, without increasing referral rates.

In summary, results from available systematic reviews trend towards a reduction in late DDH diagnoses when comparing universal US screening to selective US screening. The inconsistent findings may represent the large heterogeneity of available studies in terms of geographical locations and variation in screening strategies. Future large-scale prospective RCTs are needed to determine the comparative effectiveness of selective- or universal US screening for DDH.

However, several studies have been conducted which may provide insights into the advantages and drawbacks of selective and universal US screening.

4.3. ISSUES IN SELECTIVE US SCREENING FOR DDH

Clinical hip examinations

The foundation of traditional selective US screening is clinical hip examinations and risk factor identification.

The clinical hip examinations suffer from low reliability (9,70). While the reliability is higher in the hands of expert paediatricians or orthopaedic surgeons(69,71), the sensitivity for DDH is as low as 2.6% (72) and in one study 14% of US verified dislocated hips were incorrectly classified as reduced by expert clinical examination (8). Further, clinical examination fails to predict acetabular dysplasia in adulthood (65,66,68).

The two published RCTs comparing selective and universal US screening for DDH have both been conducted in Norway, where the primary screening is centralised to experienced paediatricians (67,73). The Norwegian RCTs did not find any significant differences in late diagnoses of DDH (>one month), and concluded that if the quality of the clinical screening was sufficiently high, the introduction of universal US did not improve the detection of DDH (74). These results should therefore be extrapolated with care to screening programs, where clinical screening is decentralised to non-expert clinical examiners.

In most countries performing selective US screening for DDH, clinical screening is not centralised to expert subspecialized physicians but rather performed by a large group of health care professionals. In Denmark the primary screeners are midwives, who perform the screening after birth, and general practitioners who perform the screening at the routine five-week follow-up (20). The positive predictive value of clinical hip examinations in detecting US verified DDH, performed by this group of screeners, is as low as 1.8% (75,76).

Additionally, in a survey of primary screeners in Denmark, 89% of respondents were able to identify written descriptions of the Ortolani manoeuvre while nearly half (41-58%) of the respondents were unable to identify the Galeazzi or hip abduction examinations, indicating a lack of knowledge among primary clinical screeners of clinical screening guidelines and recommended clinical examinations (77).

These findings may in part explain why the expansion of screening to larger groups of health care professionals lead to an increase in referral rates without a corresponding increase in treatment (78).

In study II, PFD as a referral criterion had significantly higher sensitivity in detecting DDH when compared to clinical examination. While the specificity for clinical examinations was higher and PPV was not significantly worse, clinical examination

detected only detected 15 out of 77 DDH hips while PFD detected 62 out of 77 DDH hips.

Judging from the available literature, optimal clinical screening for DDH may be achieved through centralising clinical examinations to expert examiners, which may decrease unnecessary referrals without compromising treatment.

Risk factors for DDH

Multiple risk factors for DDH have been proposed through the years which include: breech presentation, family history of DDH, multiple births, primiparity, oligohydramnios, clubfeet, congenital musculoskeletal syndromes, increased/decreased birthweight and prematurity (79–82).

In the two available meta-analyses investigating risk factors for DDH, breech presentation, female sex, primiparity and family history of DDH were found to significantly increase the risk of DDH (table 3.3) (81,82)

Risk factor analysed	**Study (n included studies)**	
	De Hundt et al 2012 (30)	Ortiz-Neira et al 2012 (31)
	OR (95% CI)	RR (95% CI)
Breech presentation	**5.7 (4.4 ; 7.4)**	**3.75 (2.3 ; 6.2)**
Female sex	**3.8 (3.0 ; 4.6)**	**2.54 (2.1 ; 3.1)**
Family history of DDH	**4.8 (2.8 ; 8.2)**	**1.4 (1.2 ; 1.6)**
Primiparity	N/A	**1.44 (1.1 ; 1.9)**
Foot deformities	3.2 (0.9 ; 12.0)	N/A
Oligohydramnios	2.5 (0.8 ; 8.2)	N/A
Mulitple births	0.5 (0.0 ; 13.6)	N/A

Table 4.3: Analysed risk factors and results in the meta-analyses by De Hundt et al 2012 and Ortiz-Neira et al 2012. Factors that were found to significantly increase the risk are marked in bold (81,82).

Although there is international variation in the risk factors used for screening, breech presentation and family history of DDH are generally accepted as referral criteria for follow-up hip US in selective US screening (18–20).

In Denmark, risk factors used as referral criteria for follow-up hip US are regionally defined and show slight variation (83). In study II of this thesis, the risk factors used were breech presentation, multiple births, oligohydramnios, clubfeet and family history of DDH. While recognition of risk factors for DDH among primary screeners is low (83), the use of risk factors as referral criteria in study II resulted in a 72% false positive rate even when including the detection of immature IIa hips.

The high false positive rate for risk factors used as dichotomous referral/no referral criteria may reflect an insufficient relative risk increase of DDH when compared to the relatively low a-priori risk of DDH in the population. A more holistic approach including multiple risk factors for DDH may therefore be needed, if false positive referrals in selective US screening are to be reduced.

The included studies in the risk factor meta-analyses use DDH according to the Graf classification as outcome. While the insignificant results may stem from a lack of correlation to DDH or power in analysis, the effect of each factor may also be too subtle to register on the Graf classification scale. For instance, in a cross-sectional study of the effect of birthweight on alpha angles, Orak et al found a significant negative correlation between increased birth weight and alpha angles on hip US for females, but was unable to detect a correlation to increased hip pathology according to the Graf classification. They concluded that birthweight was not a risk factor even though it predicted changes in acetabular inclination angles in US (84) and therefore was demonstrated as an attributing factor to acetabular dysplasia.

As this example demonstrates, individual factors may increase the risk of dysplastic changes to the acetabulum but at an insufficient level to register as an increase in Graf hip type and are therefore not a risk factor in the traditional dichotomous referral/no-referral sense. Neglecting the impact of risk attributing factors might then underestimate the risk of DDH for the child. In a study by Roposch et al from 2020, excellent discrimination was demonstrated in quantifying the risk of DDH by including multiple risk factors (female sex, family history of DDH, birthweight >4,000g, positive clinical examination) in a risk-assessment analysis model for DDH. Even though they used traditional referral criteria, this model may be expanded to other factors which increase the risk of DDH to a lesser degree, providing a more holistic view of the "total risk of DDH" for the screened child (85).

Going forward it may therefore be advisable to further investigate the use of individualised risk of DDH by pooled risk estimates based on multiple risk factors and clinical findings. Considering the impact of multiple risk factors may improve the sensitivity of selective US screening, when compared to screening relying on single

risk factors as arguably ineffective referral criteria for DDH screening. Alternatively, as shown in study II, PFD screening could be a valid alternative to risk factor screening, as PFD as referral criterion outperforms risk factor screening in terms of sensitivity, specificity and PPV in the detection of DDH hips.

4.4. THE POTENTIAL ROLE OF PFD IN DDH SCREENING

Selective US screening for DDH faces multiple issues including low early detection rates, low sensitivity of clinical examinations and risk factors, and possibly higher late detection rates.

In contrast, universal Graf US screening, while outperforming selective US screening on all these parameters, is more expensive (86) and the Graf method requires extensive training of examiners to be performed correctly (24) while still suffering from low reliability (23). Conversely, universal Graf US screening may lead to overdiagnosis and overtreatment (69), but as previously discussed, overdiagnosis may presently be unavoidable if late presenting DDH is to be eradicated. Overtreatment of DDH may lead to an increase in complications, but with advances in harness treatment and increased awareness the incidence of treatment complications is low (87).

Selective US screening using the PFD method as referral criteria for Graf US may be the compromise between selective US screening relying on clinical examinations and risk factors and universal Graf US screening. As demonstrated in study I, the PFD method can be swiftly taught to novice users to a level of reliability exceeding that of the Graf method (23). In the 1-year study period of study II, detection rates increased by 72% for immature type IIa hips and 60% for type IIc hips. To expand on this, in study II only 1 type IIc hip, in the population referred for follow-up hip US, would have been missed if PFD had been the sole referral criteria as opposed to 4 type IIc hips missed if clinical examination and risk factors were used instead of PFD as referral criteria.

Currently, DDH and Graf US is tightly linked, and given the natural development of DDH as a condition, possible replacements for the Graf US method will have to demonstrate effectiveness in predicting both early- and long-term clinical outcomes. It is not currently known if the PFD method can predict such clinical outcomes, but as it is significantly correlated to both the Graf and Harcke/Terjesen US metrics (study III), which predict acetabular dysplastic changes in adulthood (68), PFD may, by association, similarly predict the acetabular development over time. While this thesis does not argue the replacement of the Graf method in diagnostic decision making, given the association between PFD, alpha angles and FHC and the PFD method's accessibility, it may, in time, be a candidate for replacing the more complex Graf measurements.

Based on the studies presented in this thesis, the PFD method, as a referral criterion for follow-up Graf hip US, outperforms clinical screening and risk factor identification in terms of sensitivity, specificity and PPV with similar referral rates to traditional referral criteria. PFD is correlated to gold standard US metrics for DDH and can be easily taught to novice users who perform the measurements with high reliability. The PFD as a referral criterion may therefore be a promising candidate for updating selective US screening programmes for DDH.

4.5. LIMITATIONS OF STUDIES INCLUDED IN THIS THESIS

The central limitations for each study are described in the attached manuscripts. In the following section, additional important limitations for each study are described.

Study I: Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound.

In study I, midwives were demonstrated to reliably perform PFD measurements after a short introduction with minimal disagreement when compared to expert musculoskeletal radiologists. To not cause unnecessary stress for the newborn and parent or to complicate the contralateral scan for the radiologist, the newborns only received the secondary PFD scan by the midwife if the newborn was calm. Distressed newborns will resist hip US examination, making accurate hip US difficult. As the midwives were only evaluated in the best-case scenario of a calm non-resisting newborn, this may cause a selection bias resulting in higher agreement levels when comparing midwives and radiologists.

Study II: Point of care ultrasound outperforms traditional referral criteria in selective screening for hip dysplasia.

PFD as referral criteria was compared to clinical hip examinations and risk factors currently performed by midwives at AUH. The resulting discrepancies in performance may be caused by the intense training the PFD trained midwives received as opposed to the suboptimal instruction in clinical examinations the compared midwives have received during their training. If the midwives who perform clinical examination received similar supervised training it is likely that the performance of clinical examinations as referral criteria would increase towards the level of expert examiners. However, training clinical examiners to detect the relatively rare cases of hip instability on an expert level may require hundreds of examinations as well as immediate US verification of instability to confirm true positives.

Information about the study was delivered by the midwives connected to the parents of eligible newborns at AUH and the department and personnel was informed about the study and it's aims. The study therefore introduced a heightened awareness of DDH screening in the department and we observed an increased focus on risk factors

and clinical hip examinations among primary screeners. The resulting possible performance bias may lead to better performance in the selective programme, which may explain the PPV of 17.4% for clinical examinations, which has previously been reported to be as low as 4.0% for primary clinical screeners (76) (75).

Despite extensive efforts to communicate the project, the lack of risks and the low predictive value of clinical- and risk factor screening to the parents of newborns born at AUH, the achieved inclusion rate was only 57%. As information wasn't collected on newborns who weren't included or their parents, we can only provide anecdotal accounts as to why the parents might refuse participation. The PFD US screening was initially planned to take place in a room immediately across the hall from the room where the clinical screening was performed, to minimise any inconvenience to the participating families. Due to logistical considerations in the department, this room was changed to one roughly 150 metres away, forcing parents participating in the study to relocate rather than walking a few steps across the hall. From personal accounts, this inconvenience seemed to play a significant part in the parents' participation. A true point-of-care PFD examination, performed immediately after clinical screening may have increased participation in PFD US.

Study III:

In this retrospective study, two of three radiologists, who performed the measurements obtained for this study, did not routinely report FHC during provocation. Rather they only reported the exact FHC percentage if FHC was below 50% in situ or during provocation. This resulting information bias could increase the correlation between PFD and FHC during provocation as a substantial number of normal FHC during provocation values are missing.

4.6. FUTURE PERSPECTIVES

4.6.1. FUTURE PERSPECTIVES IN PFD RESEARCH

To date, only one study has reported on the incidence of late diagnoses in a PFD screening programme (29), future large scale prospective studies are therefore needed to evaluate the impact of selective PFD screening on the incidence of late DDH.

In the original description of the Graf method, the child should be placed in the lateral examination position (10). But in published studies, the examination position varies between supine and lateral. This inconsistency is also present in studies of the PFD method. While the anatomy of interest in the Graf US does not change between examination positions, the application of force in the PFD method may be dependent on the examination position of the child. Future studies should therefore investigate what, if any, effect the examination position has on the PFD measurements to determine if the PFD measurements are robust across examination positions. This knowledge will enable comparison of results across studies using different examination positions and may provide additional information about the robustness of the PFD method. Additionally, the robustness may be further studied by investigating the effect of probe mispositioning on the obtained PFD measurements, as probe mispositioning significantly affects obtained Graf US images (88).

4.6.2. FUTURE PERSPECTIVES FOR RESEARCH IN DDH SCREENING

Spontaneous resolution of previously diagnosed dysplastic hips is common. Predicting the developmental potential of dysplastic hips should therefore be a key focus of future research efforts, as up 95% of follow-up examinations of immature hips may be unnecessary, representing a drain on health-care resources, especially in screening programmes with high rates of early detection, and causing unnecessary worry for the parents and children. If any clues to the developmental potential of the hip is in the standard plane Graf US images, convolutional neural networks, a subset of artificial intelligence (AI), may be able to distinguish between hips that will spontaneously resolve and those that will have persisting acetabular dysplasia thus needing follow-up and treatment. If spontaneous resolvers among dysplastic hips can be reliably identified, it may lead to a rethinking of what constitutes pathological dysplasia that needs monitoring and what constitutes immaturity which can safely be disregarded.

It may be, that the 40-year old Graf method, that relies on a single 2D representation of the acetabular surface, will be replaced. Current candidates include AI enhanced diagnostics such as the already commercially available MEDO-Hip AI software, which today can provide interpretation of the quality of obtained US images, perform the required measurements but also obtain the correct image from US sweeps of the hip (89). The reliability of AI enhanced US is still questionable, but refinement of

these systems might reduce the need of specialised US examiners, thus financially enabling larger US screening programmes. While the acetabular alpha angles are correlated to radiographic markers of dysplasia in adulthood (68), 3D hip US, which includes a complete mapping of the acetabular surface, may be the next step towards a more comprehensive understanding of acetabular dysplasia, while providing higher reliability than its 2D counterpart (90,91).

LITTERATURE LIST

1. Phelps A. Congenital dislocation of the hip. J Bone Jt Surg Am. 1891;1:132–47.

2. Klisic P. Congenital dislocation of the hip--a misleading term: brief report. J Bone Joint Surg Br [Internet]. 1989 Jan;71-B(1):136–136. Available from: https://online.boneandjoint.org.uk/doi/10.1302/0301-620X.71B1.2914985

3. Bowen JR, Kotzias-Neto A. Developmental Dysplasia of the Hip. 1st ed. Data Trace Medical Publishing; 2006. 324 p.

4. Sewell MD, Rosendahl K, Eastwood DM. Developmental dysplasia of the hip. BMJ. 2009;339(7732):1242–8.

5. Severin E. [The frequency of congenital hip dislocation and congenital equinovarus in Sweden]. Nord Med [Internet]. 1956 Feb 16;55(7):221–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/13309785

6. Polívka D. [Incidence of dysplasia of the hip joint in Czechoslovakia (author's transl)]. Acta Chir Orthop Traumatol Cech [Internet]. 1973 Dec;40(6):507–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4793964

7. Loder RT, Skopelja EN. The Epidemiology and Demographics of Hip Dysplasia. ISRN Orthop [Internet]. 2011;2011:1–46. Available from: http://www.hindawi.com/journals/isrn/2011/238607/

8. Harper P, Joseph BM, Clarke NMP, Herrera-Soto J, Sankar WN, Schaeffer EK, et al. Even Experts Can Be Fooled: Reliability of Clinical Examination for Diagnosing Hip Dislocations in Newborns. J Pediatr Orthop. 2020 Sep 1;40(8):408–12.

9. El-Shazly M, Trainor B, Kernohan WG, Turner I, Haugh PE, Johnston AF, et al. Reliability of the Barlow and Ortolani Tests for Neonatal Hip Instability. J Med Screen [Internet]. 1994 Jul 23 [cited 2020 Jan 20];1(3):165–8. Available from: http://journals.sagepub.com/doi/10.1177/096914139400100306

10. Graf R. New Possibilities for the Diagnosis of Congenital Hip Joint Dislocation by Ultrasonography. Journal of Pediatric Orthopedics; 1983.

11. Graf R. Hip Sonography. Diagnosis and Management of Infant Hip Dysplasia. 2nd ed. Springer. Springer; 2006.

12. Kutlu A, Memik R, Mutlu M, Kutlu R, Arslan A. Congenital dislocation of the hip and its relation to swaddling used in Turkey. J Pediatr Orthop [Internet]. 1992;12(5):598–602. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1517418

13. Harsanyi S, Zamborsky R, Kokavec M, Danisovic L. Genetics of developmental dysplasia of the hip. Eur J Med Genet [Internet]. 2020 Sep;63(9):103990. Available from: https://linkinghub.elsevier.com/retrieve/pii/S176972122030495X

14. Biedermann R, Riccabona J, Giesinger JM, Brunner A, Liebensteiner M, Wansch J, et al. Results of universal ultrasound screening for developmental dysplasia of the hip: a prospective follow-up of 28 092 consecutive infants. Bone Joint J. 2018 Oct;100-B(10):1399–404.

15. Staheli LT. Fundamentals of Pediatric Orthopedics. LWW; 2015.

16. Atalar H, Sayli U, Yavuz OY, Uraş I, Dogruel H. Indicators of successful use of the Pavlik harness in infants with developmental dysplasia of the hip. Int Orthop. 2007;31(2):145–50.

17. Holman J, Carroll KL, Murray KA, Macleod LM, Roach JW. Long-term follow-up of open reduction surgery for developmental dislocation of the hip. J Pediatr Orthop [Internet]. 2012 Mar;32(2):121–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22327444

18. American Academy of Orthopaedic Surgeons; American Academy of Pediactrics. Detection and Nonoperative Management of Pediatric Developmental Dysplasia of the Hip in Infants Up To Six Months of Age Evidence-Based Clinical Practice Guideline. Aaos Cpg. 2014;389.

19. NHS. NIPE guidelines on hip screening [Internet]. 2021 [cited 2023 Jan 24]. Available from: https://www.gov.uk/government/publications/newborn-and-infant-physical-examination-programme-handbook/newborn-and-infant-physical-examination-screening-programme-handbook#screening-examination-of-the-hips

20. Sundhedsstyrelsen. Anbefalinger for Svangreomsorgen 2022 [Internet]. 4th ed. https://www.sst.dk/-/media/Udgivelser/2021/Anbefalinger-svangreomsorgen/Svangreomsorg-2022-ny.ashx?sc_lang=da&hash=F89081C3D9BCBF3367F0098F1961FF89. 2022. 316 p. Available from: https://www.sst.dk/-/media/Udgivelser/2021/Anbefalinger-svangreomsorgen/Svangreomsorg-2022-

ny.ashx?sc_lang=da&hash=F89081C3D9BCBF3367F0098F1961FF89

21. Ulziibat M, Munkhuu B, Schmid R, Baumann T, Essig S. Implementation of a nationwide universal ultrasound screening programme for developmental dysplasia of the neonatal hip in mongolia. J Child Orthop. 2020;14(4):273–80.

22. O'Beirne JG, Chlapoutakis K, Alshryda S, Aydingoz U, Baumann T, Casini C, et al. International Interdisciplinary Consensus Meeting on the Evaluation of Developmental Dysplasia of the Hip Internationales interdisziplinares Konsensustreffen zur Evaluation der Diagnostik und Therapie der angeborenen Hüftdysplasie Authors. Ultraschall Med [Internet]. 2019;40(4):454–64. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31195424#

23. Quader N, Schaeffer EK, Hodgson AJ, Abugharbieh R, Mulpuri K. A Systematic Review and Meta-analysis on the Reproducibility of Ultrasound-based Metrics for Assessing Developmental Dysplasia of the Hip. J Pediatr Orthop [Internet]. 2018 Jul;38(6):e305–11. Available from: https://journals.lww.com/01241398-201807000-00002

24. Ömeroğlu H. Significant Assessment Discrepancies Occur Between the Examiners Having and Not Having Further Special Hands-on Course Trainings in Infantile Hip Ultrasonography by the Graf Method. Indian J Orthop [Internet]. 2021 Sep 13; Available from: https://link.springer.com/10.1007/s43465-021-00514-2

25. Harcke HT, Clarke NM, Lee MS, Borns PF, MacEwen GD. Examination of the infant hip with real-time ultrasonography. J Ultrasound Med [Internet]. 1984;3(3):131–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6726860

26. Morin C, Harcke HT, MacEwen GD. The infant hip: real-time US assessment of acetabular development. Radiology [Internet]. 1985 Dec;157(3):673–7. Available from: http://pubs.rsna.org/doi/10.1148/radiology.157.3.3903854

27. Terjesen T, Runden T, Tangerud Å. Ultrasonography and radiography of the hip in infants. Acta Orthop. 1989;60(6):651–60.

28. Couture A, Baud C, Prodhomme O, Saquintaah M, Veyrac C. Ultrasound of the neonatal hip: initial evaluation and follow-up. J Radiol [Internet]. 2011;92(2):142–65. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21352746

29. Tréguier C, Chapuis M, Branger B, Bruneau B, Grellier A, Chouklati K, et al. Pubo-femoral distance: An easy sonographic screening test to avoid late diagnosis of developmental dysplasia of the hip. Eur Radiol. 2013;23(3):836–44.

30. Teixeira SR, Dalto VF, Maranho DA, Zoghbi-Neto OS, Volpon JB, Nogueira-Barbosa MH. Comparison between Graf method and pubo-femoral distance in neutral and flexion positions to diagnose developmental dysplasia of the hip. Eur J Radiol [Internet]. 2015;84(2):301–6. Available from: http://dx.doi.org/10.1016/j.ejrad.2014.11.003

31. Ban Y, Luan Q, Shi M, Sun B, Li T, Wang Y, et al. Establishing reference values for the pubofemoral distance in normal infant medial hips by ultrasound. Acta radiol. 2020;(324).

32. Miller GE. The assessment of clinical skills/competence/performance. Acad Med [Internet]. 1990 Sep;65(9 Suppl):S63-7. Available from: http://journals.lww.com/00001888-199009000-00045

33. Costa-Santos C, Bernardes J, Ayres-De-Campos D, Costa A, Costa C. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. J Clin Epidemiol. 2011;64(3):264–9.

34. Liu X. Classification accuracy and cut pointselection. Stat Med. 2012;31(23):2676–86.

35. Smits N. (Smits 2010) A note on Youden's J and its cost ratio. 2010;2–5.

36. Husum H-C, Hellfritzsch MBMB, Hardgrib N, Møller-Madsen B, Rahbek O. Suggestion for new 4.4 mm pubo-femoral distance cut-off value for hip instability in lateral position during DDH screening. Acta Orthop [Internet]. 2019;3674(1):88–93. Available from: https://www.tandfonline.com/doi/full/10.1080/17453674.2018.1554404

37. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang [Internet]. 2013 Apr;68(3). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24403724

38. Efron B. Bootstrap Methods: Another Look at the Jackknife. Ann Stat [Internet]. 1979 Jan 1;7(1):121–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3719049

39. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol [Internet]. 2011 Jun;64(6):701–2. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0895435610004336

40. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions [Internet]. 3rd ed. Wiley; 2003. (Wiley Series in Probability and Statistics). Available from: https://onlinelibrary.wiley.com/doi/book/10.1002/0471445428

41. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics [Internet]. 1977 Mar;33(1):159. Available from: https://www.jstor.org/stable/2529310?origin=crossref

42. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033–9.

43. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull [Internet]. 1979;86(2):420–8. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.86.2.420

44. Shorter D, Hong T, Osborn DA. Cochrane Review: Screening programmes for developmental dysplasia of the hip in newborn infants. Evidence-Based Child Heal. 2013;8(1):11–54.

45. Woolacott NF, Puhan MA, Steurer J, Kleijnen J. Ultrasonography in screening for developmental dysplasia of the hip in newborns: Systematic review. Br Med J. 2005;330(7505):1413–5.

46. Jung HW, Jang WY. Effectiveness of different types of ultrasonography screening for developmental dysplasia of the hip: A meta-analysis. Medicine (Baltimore). 2020;99(50):e23562.

47. Ewald E, Kiesel E. Screening for Developmental Dysplasia of the Hip: Recommendation Statement. Pediatrics [Internet]. 2006 Mar 1;117(3):898–902. Available from: https://publications.aap.org/pediatrics/article/117/3/898/68606/Screening-for-Developmental-Dysplasia-of-the-Hip

48. NHS. National Screening Committee Child Health Sub-Group Report Dysplasia of the hip September 2004. 2004.

49. Aarvold A, Perry DC, Mavrotas J, Theologis T, Katchburian M, BSCOS DDH Consensus Group. The management of developmental dysplasia of the hip in children aged under three months : a consensus study from the British Society

for Children's Orthopaedic Surgery. Bone Joint J [Internet]. 2023 Feb;105-B(2):209–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/36722054

50. Mulpuri K, Song KM, Gross RH enr., Tebor GB, Otsuka NY oshinob., Lubicky JP, et al. The American Academy of Orthopaedic Surgeons Evidence-Based Guideline on Detection and Nonoperative Management of Pediatric Developmental Dysplasia of the Hip in Infants up to Six Months of Age. J Bone Joint Surg Am. 2015;97(20):1717–8.

51. Roovers EA, Boere-Boonekamp MM, Mostert AK, Castelein RM, Zielhuis GA, Kerkhoff THM. The natural history of developmental dysplasia of the hip: Sonographic findings in infants of 1-3 months of age. J Pediatr Orthop Part B. 2005;14(5):325–30.

52. Rosendahl K, Dezateux C, Fosse KR, Aase H, Aukland SM, Reigstad H, et al. Immediate treatment versus sonographic surveillance for mild hip dysplasia in newborns. Pediatrics. 2010 Jan;125(1):e9-16.

53. Pollet V, Castelein RM, van de Sande M, Witbreuk M, Mostert AK, Besselaar A, et al. Abduction treatment in stable hip dysplasia does not alter the acetabular growth: results of a randomized clinical trial. Sci Rep. 2020;10(1):9647.

54. Sundhedsstyrelsen. Sundhedsstyrelsens retningslinier af 12. august 2004 for neonatal hørescreening. 2004;(august 2004):1–18.

55. Kuitunen I, Uimonen MM, Haapanen M, Sund R, Helenius I, Ponkilainen VT. Incidence of Neonatal Developmental Dysplasia of the Hip and Late Detection Rates Based on Screening Strategy: A Systematic Review and Meta-analysis. JAMA Netw open [Internet]. 2022;5(8):e2227638. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35980635

56. Higgins JPT. Measuring inconsistency in meta-analyses. BMJ [Internet]. 2003 Sep 6;327(7414):557–60. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.327.7414.557

57. Deeks J, Higgins J, Altman D. Cochrane Handbook - Meta-Analyses [Internet]. Cochrane Handbook for Systematic Reviews of Interventions. 2022 [cited 2022 Mar 21]. Available from: https://training.cochrane.org/handbook/current/chapter-10#section-10-11

58. Gyurkovits Z, Sohár G, Baricsa A, Németh G, Orvos H, Dubs B. Early detection of developmental dysplasia of hip by ultrasound. HIP Int [Internet].

2019 Sep 30;112070001987968. Available from: http://journals.sagepub.com/doi/10.1177/1120700019879687

59. Moosa NK, Kumar PT, Mahmoodi SM. Incidence of developmental dysplasia of the hip in Dubai. Saudi Med J [Internet]. 2009 Jul;30(7):952–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19618014

60. Kishore Kumar R, Shah P, AN R, Rajan R. Diagnosing Developmental Dysplasia of Hip in Newborns Using Clinical Screen and Ultrasound of Hips—An Indian Experience. J Trop Pediatr [Internet]. 2016 Jun;62(3):241–5. Available from: https://academic.oup.com/tropej/article-lookup/doi/10.1093/tropej/fmv107

61. Geertsema D, Meinardi JE, Kempink DRJ, Fiocco M, van de Sande MAJ. Screening program for neonates at risk for developmental dysplasia of the hip: comparing first radiographic evaluation at five months with the standard twelve week ultrasound. A prospective cross-sectional cohort study. Int Orthop [Internet]. 2019 Aug 18;43(8):1933–8. Available from: http://link.springer.com/10.1007/s00264-018-4089-2

62. Tschauner C, Klapsch W, Baumgartner A, Graf R. „Reifungskurve“ des sonographischen Alpha-Winkels nach GRAF unbehandelter Hüftgelenke im ersten Lebensjahr. Z Orthop Ihre Grenzgeb [Internet]. 2008 Mar 18;132(06):502–4. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-2008-1039477

63. Biedermann R, Eastwood DM. Universal or selective ultrasound screening for developmental dysplasia of the hip? A discussion of the key issues. J Child Orthop. 2018;12(4):296–301.

64. Broadhurst C, Rhodes AML, Harper P, Perry DC, Clarke NMP, Aarvold A. What is the incidence of late detection of developmental dysplasia of the hip in England?: a 26-year national study of children diagnosed after the age of one. Bone Joint J. 2019 Mar;101-B(3):281–7.

65. Sink EL, Ricciardi BF, Torre K Dela, Price CT. Selective ultrasound screening is inadequate to identify patients who present with symptomatic adult acetabular dysplasia. J Child Orthop. 2014;8(6):451–5.

66. Engesæter IØ, Lie SA, Lehmann TG, Furnes O, Vollset SE, Engesæter LB, et al. Neonatal hip instability and risk of total hip replacement in young adulthood: Follow-up of 2,218,596 newborns from the Medical Birth Registry of Norway in the Norwegian Arthroplasty Register. Acta Orthop. 2008;79(3):321–6.

67. Rosendahl K, Markestad T, Lie RT. Ultrasound screening for developmental dysplasia of the hip in the neonate: the effect on treatment rate and prevalence of late cases. Pediatrics [Internet]. 1994 Jul;94(1):47–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8008537

68. Laborie LB, Lie SA, Rosendahl K. Radiographic markers of hip dysplasia in young adults: predictive effect of factors in early life. BMC Musculoskelet Disord [Internet]. 2023;24(1):119. Available from: https://doi.org/10.1186/s12891-023-06199-y

69. Olsen SF, Blom HC, Rosendahl K. Introducing universal ultrasound screening for developmental dysplasia of the hip doubled the treatment rate. Acta Paediatr [Internet]. 2018 Feb;107(2):255–61. Available from: https://onlinelibrary.wiley.com/doi/10.1111/apa.14057

70. Rosendahl K, Markestad T, Lie RT. Ultrasound in the early diagnosis of congenital dislocation of the hip: The significance of hip stability versus acetabular morphology. Pediatr Radiol. 1992;22(6):430–3.

71. Krikler S, Dwyer N. Comparison of results of two approaches to hip screening in infants. J Bone Joint Surg Br [Internet]. 1992 Sep;74-B(5):701–3. Available from: https://online.boneandjoint.org.uk/doi/10.1302/0301-620X.74B5.1527116

72. Subaşı İÖ, Veizi E, Çepni Ş, Alkan H, Oğuz T, Fırat A. Clinical Examination Findings Can Accurately Diagnose Developmental Dysplasia of The Hip—A Large, Single-Center Cohort. Children [Internet]. 2023 Feb 4;10(2):304. Available from: https://www.mdpi.com/2227-9067/10/2/304

73. Holen KJ, Surgeon Tegnander OA, Resident Bredland OT, Johansen OJ, Paediatrician D Saether CO, Surgeon C, et al. Universal or selective screening of the neonatal hip using ultrasound? J Bone Jt Surg [Br]. 2002 Aug;84(6):886–90.

74. Laborie LB, Engesæter I, Lehmann TG, Eastwood DM, Engesæter LB, Rosendahl K. Screening strategies for hip dysplasia: Long-term outcome of a randomized controlled trial. Pediatrics. 2013;132(3):492–501.

75. Choudry QA, Paton RW. Neonatal screening and selective sonographic imaging in the diagnosis of developmental dysplasia of the hip. Bone Jt J. 2018;100B(6):806–10.

76. Husum H-C, Ghaffari A, Rytoft LA, Svendsson J, Harving S, Kold S, et al. Positive predictive values in clinical screening for developmental dysplasia of

the hip. Acta Paediatr [Internet]. 2021 Aug 11;110(8):2430–4. Available from: https://onlinelibrary.wiley.com/doi/10.1111/apa.15896

77. Husum HC, Maimburg RD, Kold S, Thomsen JL, Rahbek O. Self-reported knowledge of national guidelines for clinical screening for hip dysplasia: a web-based survey of midwives and GPs in Denmark. BJGP Open. 2021;5(4):1–8.

78. Düppe H, Danielsson LG. Screening of neonatal instability and of developmental dislocation of the hip. A survey of 132,601 living newborn infants between 1956 and 1999. J Bone Joint Surg Br [Internet]. 2002;84(6):878–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12211683

79. Dunn PM. Perinatal observations on the etiology of congenital dislocation of the hip. Clin Orthop Relat Res [Internet]. 1976 Sep;(119):11–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/954299

80. Chan A, McCaul KA, Cundy PJ, Haan EA, Byron-Scott R. Perinatal risk factors for developmental dysplasia of the hip. Arch Dis Child - Fetal Neonatal Ed [Internet]. 1997 Mar 1;76(2):F94–100. Available from: https://fn.bmj.com/lookup/doi/10.1136/fn.76.2.F94

81. De Hundt M, Vlemmix F, Bais JMJ, Hutton EK, De Groot CJ, Mol BWJ, et al. Risk factors for developmental dysplasia of the hip: A meta-analysis. Eur J Obstet Gynecol Reprod Biol [Internet]. 2012;165(1):8–17. Available from: http://dx.doi.org/10.1016/j.ejogrb.2012.06.030

82. Ortiz-Neira CL, Paolucci EO, Donnon T. A meta-analysis of common risk factors associated with the diagnosis of developmental dysplasia of the hip in newborns. Eur J Radiol. 2012 Mar;81(3):e344-51.

83. Husum HC, Thomsen JL, Kold S, Maimburg RD, Rahbek O. Referral criteria recognition of screeners in the Danish screening programme for hip dysplasia. Dan Med J. 2022;69(2):1–8.

84. Orak MM, Karaman O, Gursoy T, Cagirmaz T, Oltulu I, Muratli HH. Effect of increase in birth weight in a newborn on hip ultrasonography. J Pediatr Orthop B. 2015 Nov;24(6):507–10.

85. Roposch A, Protopapa E, Malaga-Shaw O, Gelfer Y, Humphries P, Ridout D, et al. Predicting developmental dysplasia of the hip in at-risk newborns. BMC Musculoskelet Disord. 2020 Jul;21(1):442.

86. Harper P, Gangadharan R, Poku D, Aarvold A. Cost Analysis of Screening Programmes for Developmental Dysplasia of the Hip: A Systematic Review. Indian J Orthop [Internet]. 2020;55(6):1402–9. Available from: https://doi.org/10.1007/s43465-021-00501-7

87. Zhi X, Xiao X, Wan Y, Wei P, Canavese F, Xu H. Tübingen hip flexion splint for the treatment of developmental dysplasia of the hip in children younger than six months age: a meta-analysis. J Child Orthop [Internet]. 2021 Aug 20;15(4):402–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34476031

88. Jaremko JL, Mabee M, Swami VG, Jamieson L, Chow K, Thompson RB. Potential for change in US diagnosis of hip dysplasia solely caused by changes in probe orientation: patterns of alpha-angle variation revealed by using three-dimensional US. Radiology. 2014 Dec;273(3):870–8.

89. Hareendranathan AR, Chahal B, Ghasseminia S, Zonoobi D, Jaremko JL. Impact of scan quality on AI assessment of hip dysplasia ultrasound. J Ultrasound [Internet]. 2022 Jun 5;25(2):145–53. Available from: https://link.springer.com/10.1007/s40477-021-00560-4

90. Quader N, Hodgson AJ, Mulpuri K, Cooper A, Abugharbieh R. A 3D femoral head coverage metric for enhanced reliability in diagnosing hip dysplasia. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2017;10433 LNCS:100–7.

91. Quader N, Hodgson AJ, Mulpuri K, Cooper A, Garbi R. 3-D Ultrasound Imaging Reliability of Measuring Dysplasia Metrics in Infants. Ultrasound Med Biol. 2021;47(1):139–53.

CHAPTER 5. APPENDICES

List of appendices

- Study I: Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound (published Sep 2022, MDPI Children)

- Study II: Point of care pubo-femoral distance ultrasound outperforms traditional referral criteria in selective screening for hip dysplasia (submitted for publication)

- Study III: Pubo femoral distances correlate to acetabular morphology and sonographic instability in screening for hip dysplasia. (Submitted for publication)

5.1. APPENDIX I

The following paper has been adapted from https://doi.org/10.3390/children9091345 to follow the formatting for PhD theses dictated by the clinical institute, Aalborg university. The original paper can be accessed through the above link.

Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound

Hans-Christen Husum 1,2 , Michel Bach Hellfritzsch 2,3 , Rikke Damkjær Maimburg 2,4 , Mads Henriksen 2,3, Natallia Lapitskaya 2,3, Bjarne Møller-Madsen 2,5 and Ole Rahbek 1,2

1 Interdisciplinary Orthopaedics, Aalborg University Hospital, 9000 Aalborg, Denmark

2 Danish Paedatric Orthopaedic Research, Aarhus University Hospital, 8200 Aarhus, Denmark

3 Department of Radiology, Aarhus University Hospital, 8200 Aarhus, Denmark

4 Department of Obstetrics and Gynecology, Aarhus University Hospital, 8200 Aarhus, Denmark

5 Department of Children's Orthopaedics, Aarhus University Hospital, 8200 Aarhus, Denmark

Abstract

The pubo-femoral distance (PFD) has been suggested as an ultrasound screening tool for developmental dysplasia of the hip (DDH). The aim of this study was to examine if midwives undergoing minimal training could reliably perform pediatric hip ultrasound and PFD measurements. Eight recruited midwives performed two rounds of independent blinded PFD measurements on 15 static ultrasound images and participated in four supervised live-scanning sessions. The midwives were compared to a group of three experienced musculoskeletal radiologists. Reliability was evaluated using inter-rater correlation coefficients (ICC). Linear regression was used to quantify the learning curve of the midwives as a group. There was near complete intra- and inter-rater agreement (ICC > 0.89) on static ultrasound images across both rounds of rating for midwives and radiologists. The midwives performed a mean of 29 live hip scans (range 24–35). The mean difference between midwives and supervising radiologists was 0.36 mm, 95% CI (0.12–0.61) for the first session, which decreased to 0.20 mm, 95% CI (0.04–0.37) in the fourth session. ICC for PFD measurements increased from 0.59 mm, 95% CI (0.37–0.75) to 0.78 mm, 95% CI (0.66–0.86) with progression in sessions. We conclude that midwives reliably perform PFD measurements of pediatric hips with minimal training.

1. Introduction

Developmental dysplasia of the hip (DDH) is a condition of underdevelopment of the hip joint and ranges from mild acetabular shallowing to complete hip dislocation. With an incidence of 0.8%, it is the most common musculoskeletal disorder in children [1]. Treatment of DDH is time-sensitive as early diagnosed cases can be treated conservatively and successfully with hip bracing [2], and delayed diagnosis worsens the prognosis and increases the necessity of surgery and risk of complications [3]. Because of the relatively high incidence, and the time-sensitive nature of the condition, ultrasound screening has been widely implemented in high-resource countries. The predominant screening strategy is selective ultrasound screening based on the Graf ultrasound method [4]. At-risk newborns receive a hip ultrasound based on clinical examinations of the hip or the presence of risk factors for DDH. In contrast, in universal ultrasound screening every child receives a hip ultrasound regardless of clinical findings and/or risk factors. The only Cochrane review on DDH screening to date was inconclusive in whether to recommend selective- or universal ultrasound screening [5] due to a lack of decisive evidence. Since the Cochrane review was published, evidence increasingly points towards limitations in the selective screening approach, as 85% of patients treated for DDH do not fulfill the criteria for selective ultrasound screening [6]. Further, primary screeners have insufficient knowledge of the clinical examinations and risk factors which constitute the foundation of the selective screening approach [7–9]. These findings might explain why 26 years of

selective ultrasound screening in the UK has failed to reduce the rate of late DDH diagnoses. [10]. Since the 1990s, universal DDH ultrasound screening of all newborns with the Graf method has been performed in parts of Austria and Germany, resulting in near-eradication of late DDH diagnoses [1], and in the lowering of treatment costs although diagnostic expenses have increased similarly [11]. The international consensus on DDH screening is now shifting towards universal DDH ultrasound screening based on Graf's method [12].

A central obstacle in the implementation of universal Graf ultrasound screening is, however, the increase in diagnostic expenses and demand for specialized radiological resources in part due to the complexity of the Graf ultrasound method. In 2013 the pubo-femoral distance (PFD) was proposed as an alternative, less complex ultrasound measurement for DDH screening [13] with a high sensitivity and specificity for DDH [14]. The PFD is defined as the minimal, measurable distance between the medial femoral epiphysis and the pubic bone while applying lateralizing stress to the hip. Minimal experience is necessary to perform the measurement reliably [13]. However, the approach has only been documented when performed by radiologists. To reduce the need for specialized radiological resources and in turn the economic impact of implementing a universal PFD screening program, we hypothesized that the PFD method can be taught with minimal instruction to non-physician health-care professionals with little or no prior experience in ultrasonography while still observing a high degree of accuracy. For this purpose, midwives were selected as they are the health care professionals making the primary clinical examination of the newborn in the Danish neonatal screening program.

Our aim was firstly to demonstrate the agreement in midwife PFD measurements on static and live ultrasound images compared with measurements performed by experienced musculoskeletal radiologists. Secondly, our aim was to quantify the learning curve for the PFD ultrasound measurement among midwives.

2. Materials and Methods

2.1. Study Design

This was a prospective observational study. Reporting follows the STROBE guidelines for reporting on observational studies [15].

2.2. Participants

We recruited midwives for training in PFD measurement from the Department of Obstetrics and Gynecology at Aarhus University Hospital, Denmark. Recruited midwives completed a demographic survey which included experience as a midwife, prior experience in ultrasound examinations, and number of yearly clinical DDH screenings performed. For comparison, three musculoskeletal radiologists were

recruited from the Department of Radiology at Aarhus University Hospital, with respectively, 21, 7 and 1.5 years of experience in pediatric hip ultrasound examinations.

2.3. Constructing the PFD Training Program

The midwives' training program was conceptualized on the theoretical background described in Miller's pyramid [16] to achieve professional clinical competence. The learning content was developed on the principles of blended learning to provide both different learning styles and learning environments. We started the training program with an online introduction film about hip dysplasia followed by a traditional theoretical lecture advancing with subsequent on-site practical demonstration and exercises with online introduction

and instructional videos.

2.4. Lectures and Workshops

Each midwife was instructed to watch a 10-min online video introduction on DDH and PFD screening prior to participating in a two-hour theoretical group lecture on basic anatomy, pathogenesis and treatment of DDH as well as an introduction to pediatric hip ultrasound, with an emphasis on the PFD measurements and video demonstrations of how to obtain it. Afterwards, each midwife participated in two workshops. In the first workshop, the midwives were evaluated in a best-case scenario, where they were presented with bestpractice static hip ultrasound images in the standard projection according to Graf without annotations, obtained by a senior musculoskeletal radiologist with 21 years of experience. Each midwife performed PFD measurements on 15 images using Picture Archiving and Communication Software (PACS) (Impax, client 6.5 AGFA Healthcare, Mortsel, Belgium). Seven days later, in the second workshop, the participants repeated these measurements on the same images. For comparison, the recruited radiologists performed the same measurements with seven-day intervals. The measurement exercises of the midwives were monitored by the first author, and no instructions other than technical support were provided. All raters were blinded to the measurements performed previously by themselves and others and they were instructed not to share information on measurements.

2.5. Supervised Live-Scanning Sessions

After completing the lecture and workshops, each midwife received a 30-min introduction to the MINDRAY TE7 (Mindray Medical International, Shenzhen, China) ultrasound scanner as well as a brief introduction to general sonography. Each midwife then participated in four sessions of supervised live scans of pediatric hips over the course of two weeks as part of the DDH screening program at our institution

where PFD measurements are routinely measured. Scanning sessions were supervised by one of the three senior musculoskeletal radiologists and integrated into the current ultrasound screening program for DDH in the Radiological Outpatient Clinic at our institution. The live-scanning sessions took place between September 2021 and December 2021. For the purpose of this study, a separate MINDRAY TE7 ultrasound scanner with a simplified user-interface and high frequency (16 MHz) linear transducer was acquired and calibrated specifically for pediatric hip ultrasound. In accordance with the institution protocol, the radiologist performed a hip ultrasound on one side with the newborn in the lateral decubitus position according to the method described by Graf, Tréguier and Couture [4,13,17] using a 10 MHz linear transducer (Model: Canon Aplio i800; Canon Medical Systems, Tokyo, Japan). The midwife then repeated the scan using the MINDRAY TE7 ultrasound scanner and measured the PFD. Finally, to avoid bias introduced by using two different scanners, the radiologist would perform the PFD measurement using the MINDRAY TE7 scanner. The entire sequence was then repeated for the opposite hip.

The criteria for the ultrasound scan were initially visualization of the femoral head and the lateral epiphysis of the pubic bone while adducting the knee and performing a Barlow equivalent lateralizing stress maneuver on the hip joint to perform the PFD measurement. As the midwife gained experience, to ensure consistency and repeatability, the criteria were expanded to include the horizontal plane of ilium and the bony and cartilaginous acetabular roof in accordance with the method described by Tréguier and Couture [13,17] (Figure 1).

As these sessions also functioned as a training regimen in pediatric hip ultrasound for the midwives, the radiologists were free to instruct the midwives and give feedback as needed until the midwife had performed the PFD measurement. Neither the midwife nor the radiologist was blinded to the measurement results but they did not make repeat PFD measurements of the same hip and were instructed to disregard the measurements of the other party. The first author was present to enforce these instructions.

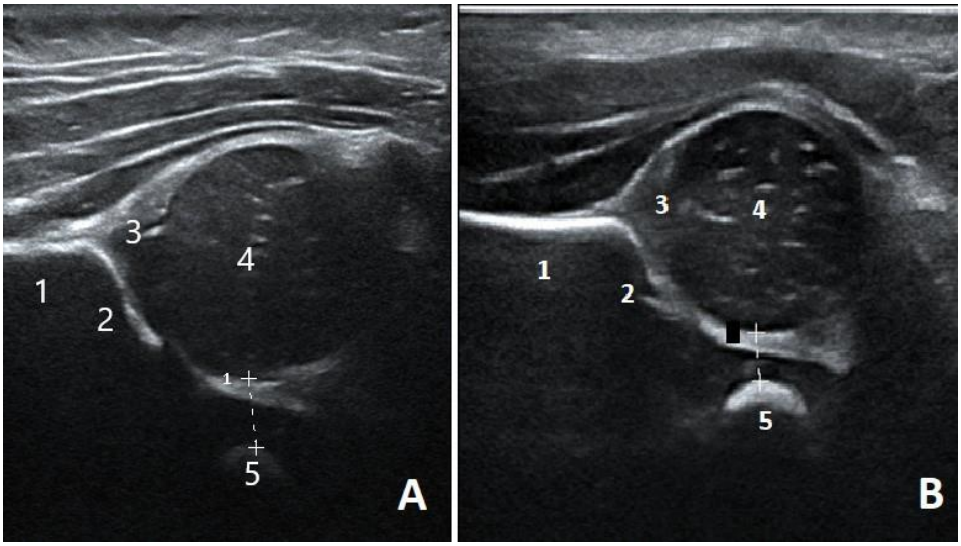


Figure 1. Ultrasound image of two different newborn hips obtained by a musculoskeletal radiologist (**A**) and a midwife (**B**) depicting the quality criteria for the PFD measurement: A horizontal ilium (1), the bony (2) and cartilaginous (3) acetabular roof, the femoral head (4) and the lateral epiphysis of the pubic bone (5). The PFD is the minimal distance between the medial femoral epiphysis and the pubic bone (dotted line). PFD = Pubo-femoral distance.

2.6. Data Sources

Static hip ultrasound images without annotations for the PFD measurement workshop were acquired from the production of the existing ultrasound screening program at our institution, and all were performed by a senior musculoskeletal radiologist according to the methods of Graf, Tréguier and Couture [4,13,17]. All measurements from the workshops and live-scanning sessions were anonymized and registered directly by the first author in a General Data Protection Regulation compliant REDCap database.

2.7. Statistical Methods

2.7.1. Workshop Measurements

We analyzed intra- and inter-rater reliability of PFD measurements within and between the group of radiologists and midwives using intraclass correlation coefficients (ICC) with accompanying 95% confidence intervals (CI). Intra-rater ICCs were calculated as twoway mixed effects, single measurement with absolute agreement, while two-way random effect was used for the inter-rater ICCs. We interpreted ICCs according to Portney and Watkins [18] with a value of 0, 0.75 and 1 representing no agreement, good agreement and complete agreement, respectively.

2.7.2. Supervised Live-Scanning Sessions

To investigate any correlation in agreement with increasing PFD values, and to illustrate the overall progress in agreement between midwives and radiologists with increasing measurement experience among the midwives, a Bland–Altman (BA) plot for each session was made with mean difference and limits of agreement (LOA). A scatter plot was made of the absolute differences in PFD values between radiologists and midwives as a function of improved experience among the midwives, defined as the cumulative number of patients scanned. The scatter plot included the mean absolute differences for the midwives as a group, as well as a fitted linear regression with 95% CI. To quantify the agreement as the midwives gained experience, a linear regression was performed with absolute difference as the dependent variable and number of patients scanned per midwife as the explanatory variable while controlling for previous ultrasound experience as a dichotomous yes/no variable. Normal distribution of PFD differences was inspected using QQ plots and a significance level of 5% was applied. Statistical analyses were performed using Stata version 16.1 (StataCorp, College Station, TX, USA).

2.8. Ethics

This was a quality control study which followed the routine DDH screening program at our institution. Findings had no impact on patient treatment or diagnosis. As per the guidelines from the Danish National Center for Ethics, ethical approval and written consent was not needed.

3. Results

Eight midwives were included in this study. Mean years of seniority as a midwife was 11 years (range 4–27 years), mean number of yearly clinical DDH screening was 68 (range 10–110). Two midwives performed fetal ultrasound in clinical practice and had done so for one year (Table 1).

Years of seniority as a midwife (mean (range))	11 years (4 years–27 years)
Number of clinical DDH screenings (Ortolani and Barlow maneuvers) yearly (mean (range))	68 (10–110)
Uses ultrasound in clinical practice (fetal ultrasound) (yes/no)	2/6
Years of experience using ultrasound in clinical practice (n = 2) (range)	1 (1)

Table 1. Demographics of recruited midwives.

3.1. Workshops

There was near complete intra- and inter-rater agreement (ICC > 0.89) in PFD measurements of static ultrasound images in both workshops across and within the midwife and radiologist groups. No difference was found in ICC between the groups of midwives and radiologists or across rounds of rating (Table 2).

	Radiologists (n=3)	**Midwives (n=9)**
Interrater RAD/MW ICC (95% CI) workshop 1	0.99 [0.86;0.99]	
Interrater RAD/MW ICC (95% CI) workshop 2	0.99 [0.92;0.99]	
Interrater ICC (95% CI) within group workshop 1	0.93 [0.84;0.97]	0.89 [0.80;0.95]
Interrater ICC (95% CI) within group workshop 2	0.95 [0.83;0.98]	0.95 [0.90;0.98]
Intrarater ICC (95% CI) group average between workshops	0.98 [0.93;0.99]	0.99 [0.84;0.99]

Table 2. Inter-rater and intra-rater ICCs of PFD measurements made on 15 static pediatric hip ultrasound images across two rating workshops. ICC = Inter-rater correlation coefficient, PFD = Pubo-femoral distance, RAD = Musculoskeletal radiologist, MW = Midwife.

3.2. Supervised Live-Scanning Sessions

The mean number of hips scanned by the midwives over the four supervised livescanning sessions was 29 (range 24–35) with a total of 237 hips scanned. Inspection of BA plots did not reveal any dependency between differences in PFD and mean PFD values. In the first supervised session the mean difference was 0.36 mm (LOA −1.42; 2.14 mm) which decreased to 0.2 mm (LOA −1.04; 1.44 mm) in the fourth and final supervised session (Figure 2).

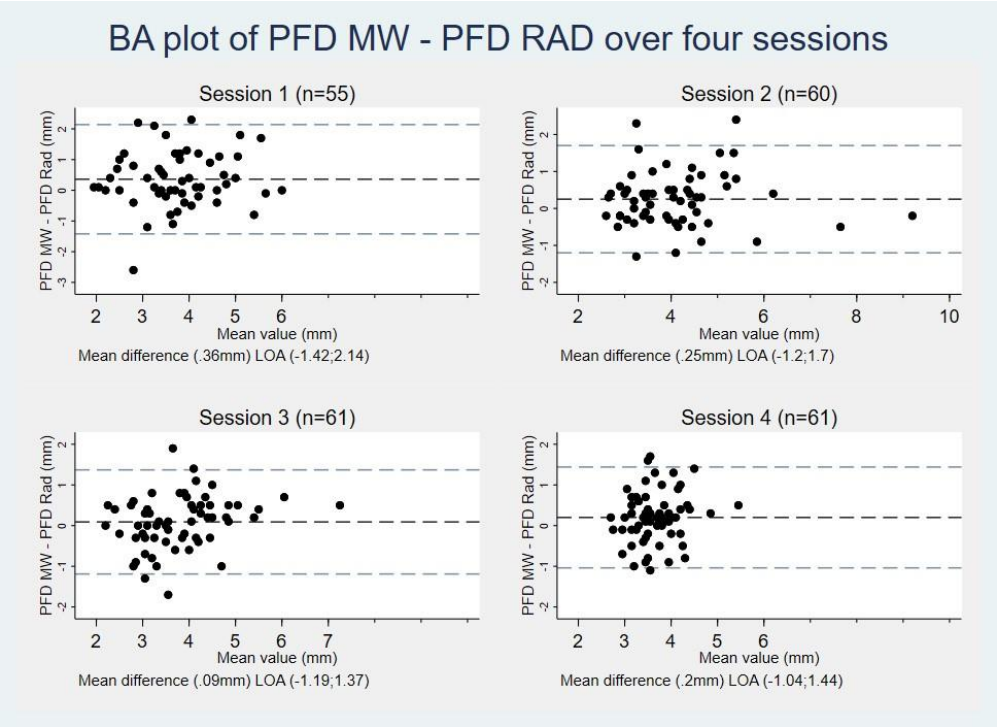


Figure 2. Bland–Altman plot of differences in PFD measurements between radiologists and midwives across four sessions of supervised pediatric hip scans. PFD = Pubo-femoral distance, MW = Midwife, RAD = Radiologist, LOA = Limits of agreement.

Inter-rater agreement between midwives and radiologists increased from ICC = 0.59 with a 95% CI (0.37; 0.75) in the first supervised session to ICC = 0.78 (0.66; 0.86) in the third supervised session. ICC for the fourth supervised session could not be evaluated due to low variance in observed values (Table 3).

Session of live scans (n=hips)	**ICC (95 % CI)**
Session 1 (n=55)	0.59 (0.37;0.75)
Session 2 (n=60)	0.81 (0.68;0.88)
Session 3 (n=61)	0.78 (0.66;0.86)
Session 4 (n=61)	0.42 (0.19;0.60)

Table 3. Inter-rater ICCs of PFD measurements made by recruited midwives and supervising radiologists across four sessions of live scans. ICC = Inter-rater correlation coefficient.

* Due to an insufficient variation in measured values, the calculated ICC value falsely quantified reliability in session 4 as low. This does not reflect the high level of agreement between the two groups of raters as can be seen from the BA plot and limits of agreement (Figure 2). The underestimation of ICC values due to low variation of observations is a known limitation of the ICC method [19], the value is only presented here for transparency.

Scatter plot inspection revealed a decrease in the range of absolute PFD differences between midwives and radiologists as the midwives became more experienced (Figure 3). Initial mean absolute PFD difference between midwives and radiologists was 0.73 mm, which decreased by 0.1 mm (95% CI 0.02–0.17 mm) for every ten scans the midwives gained in experience ($p < 0.008$) and was not associated with previous ultrasound experience of the midwife ($p = 0.51$).

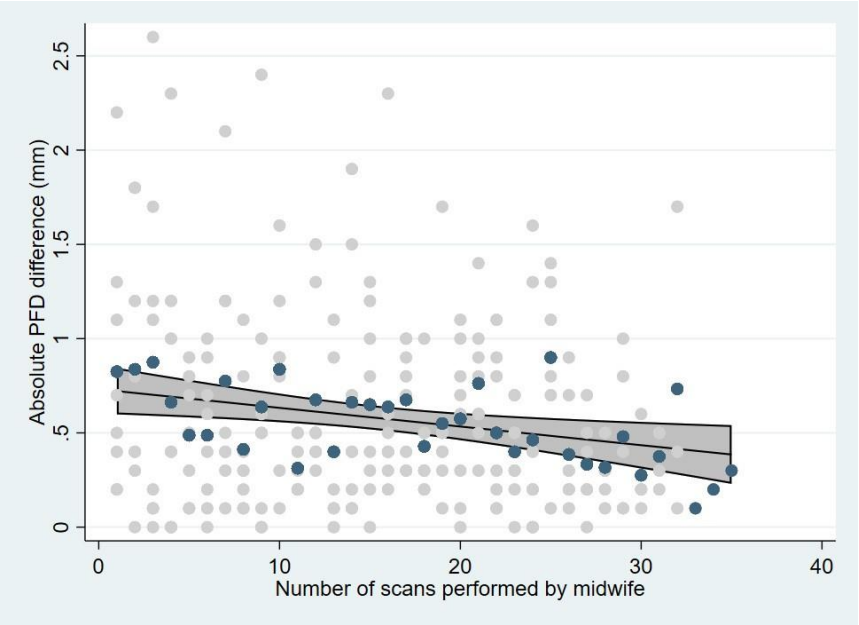


Figure 3. Scatter plot and fitted linear regression with 95% CI of absolute differences in PFD measurements between radiologists and midwives as a function of increasing scan experience. Light grey dots = individual values, blue dots = average values. PFD = Pubo-femoral distance.

4. Discussion

In this study, midwives' performance in PFD measurements on static ultrasound images showed the same level of reliability as measurements performed by senior musculoskeletal radiologists, following minimal instruction of midwives in PFD measurement by ultrasound. After a short learning program, including three supervised sessions, midwives reliably performed pediatric hip ultrasound and PFD measurements with clinically

insignificant differences when compared to experienced radiologists.

4.1. Limitations

The results may be affected by observer- and performance bias, as both groups were aware that their measurements would be logged, and the midwives, although instructed to disregard the radiologists' measurements, were not blinded to these while doing live scans. These biases may result in an observed higher agreement in observations. Due to low variance in observed values in session four, ICC values could not be calculated, which is a known statistical limitation of the ICC method [19]. However, the Bland–Altman plot depicts larger agreement in PFD

measurements between groups with increasing session numbers including session four, which is also reflected in the calculated decrease in absolute PFD difference between groups as the midwives became more experienced.

4.2. Interpretation

It is of vital importance for imaging-based screening protocols to be reliable and accurate. To date, two studies have examined the reliability of the PFD method. Teixeira et al. compared PFD measurements of one senior radiologist to one resident radiologist and found a high degree of inter-rater agreement in measurement values (ICC = 0.85) [20]. Tréguier et al. compared one senior radiologist to a radiologist in training and documented a mean measurement difference of 0.12 mm and found a substantial agreement in categorizing hips with a threshold of 6 mm (Kappa = 0.795) [13]. The present study, is the first to compare examiners with no previous experience in hip ultrasound to a group of experienced musculoskeletal radiologists, and we found similarly high levels of agreement. It is worth noting that mean measurement difference is a poor estimate for agreement, as a low value could mean that the measurements are evenly under- and overestimated. In contrast, absolute differences give a more precise estimate of agreement as they are not affected by over- and underestimations and should be used when reporting on agreement in measurements.

In a recent meta-analysis of ultrasound measurements used in DDH diagnostics, considerable variation was found for Graf's alfa and beta angle, as well as Terjesen's Femoral Head Coverage (FHC). Reliability was poor to moderate for all ultrasound-based metrics, and although variation was lowest for the alfa and beta angles, intra- and interrater agreement still varied from an ICC of 0.02 to 0.453. Further, reliability of the Graf classification of hips in the included studies was poor to moderate (Kappa 0.1–0.6) [21]. The relatively high complexity of the Graf measurements and the susceptibility to interpretation errors due to mispositioning of the ultrasound probe may lead to misinterpretation of measurements and consequently misclassification of healthy hips as dysplastic and vice versa. Jaremko et al. were able to produce clinically acceptable hip ultrasound images with a 20 degree tilt error in probe positioning, causing an alpha angle variation of 18 degrees (52–70 degrees) leading to misclassification in 54% of the hips scanned [22]. Orak et al. investigated the reliability of the alfa and beta angles and found poor inter-rater agreement among four raters with experience from >500 hip ultrasound examinations [23]. Despite the extensive experience of the raters and a pre-study consensus meeting, images from this study depicting four different interpretations of the same hip, showed inconsistencies in the application of the Graf method, translating to alpha angles ranging from 57–72 degrees or normal to mild hip dysplasia. In contrast, the PFD method is a simple single distance measurement between two distinct landmarks. Although the influence of probe positioning on PFD measurements has not yet been established, the reported high levels of inter-rater agreement in this study, among

users with varying hip ultrasound experience, suggest a higher tolerance for tilt and rotational errors of the probe.

A key issue in implementing universal ultrasound screening for DDH is the increase in diagnostic expenses and an increased demand for experienced ultrasound examiners. This is largely due to the experience needed to perform and interpret pediatric hip ultrasound using the Graf method [11]. The PFD method has been suggested as an alternative to the Graf method as a highly sensitive measurement for DDH [13,14]. As demonstrated by the present study, PFD screening measurements performed by novices did not result in lowering accuracy, thus demonstrating a possible cost-effective alternative to current screening practices, and supporting the feasibility of employing midwives to perform PFD ultrasound screening for DDH. Further studies on PFD ultrasound screening programs using non-radiologist examiners are, however, needed to evaluate the efficacy in detecting DDH through PFD screening, based on trained novice ultrasound users.

Generalizability

The degree of dysplasia based on Graf's method was not registered for the infants scanned in this study. However, as we consider a PFD > 6.0 mm to be indicative of DDH [13], and we did not find an increase in PFD between raters with increasing mean PFD values, we believe these results are valid in a dysplastic population. The midwives in this study were diverse in terms of years of seniority, number of yearly clinical DDH screenings and previous ultrasound experience, with the majority having no experience in performing ultrasound examinations. We therefore expect these results to be applicable to health care professionals with no experience in ultrasound.

5. Conclusions

We conclude that midwives, undergoing theoretical and limited supervised practical training, are able to perform PFD measurements of pediatric hips with the same level of reliability and precision as senior musculoskeletal radiologists.

Author Contributions: Conceptualization, H.-C.H., M.B.H., R.D.M., B.M.-M. and O.R.; methodology, H.-C.H., R.D.M. and O.R.; software, H.-C.H.; validation, H.-C.H., M.B.H., M.H. and N.L.; formal analysis, H.-C.H.; investigation, M.B.H., M.H. and N.L.; data curation, M.B.H., M.H., N.L. and H.-C.H.; writing—original draft preparation, H.-C.H.; writing—review and editing, M.B.H., R.D.M.,

M.H., N.L., B.M.-M. and O.R.; visualization, H.-C.H.; supervision, O.R.; project administration, H.C.H., R.D.M., M.B.H.; funding acquisition, H.-C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Independent Research Fund Denmark Grant no. 1030-00366B.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the guidelines from the Danish National Center for Ethics. (https://nationaltcenterforetik.dk/ anmeld-forsoeg/andre-sundhedsvidenskabelige-forsoeg/naar-du-anmelder/hvad-skal-jeg-anmelde accessed 2 August 2022).

Informed Consent Statement: Patient consent was waived for this quality control study per the guidelines from the Danish National Center of Ethics.

Data Availability Statement: A notification on the data set and the possibility of gaining access will be shared to Aalborg University's research portal (VBN.aau.dk).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Biedermann, R.; Riccabona, J.; Giesinger, J.M.; Brunner, A.; Liebensteiner, M.; Wansch, J.; Dammerer, D.; Nogler, M. Results of Universal Ultrasound Screening for Developmental Dysplasia of the Hip. *Bone Jt. J.* **2018**, *100*, 1399–1404. [CrossRef] [PubMed]
2. Atalar, H.; Sayli, U.; Yavuz, O.Y.; Uras¸, I.; Dogruel, H. Indicators of Successful Use of the Pavlik Harness in Infants with Developmental Dysplasia of the Hip. *Int. Orthop.* **2007**, *31*, 145–150. [CrossRef]
3. Furnes, O.; Lie, S.A.; Espehaug, B.; Vollset, S.E.; Engesaeter, L.B.; Havelin, L.I. Hip Disease and the Prognosis of Total Hip Replacements. *J. Bone Jt. Surg.-Ser. B* **2001**, *83*, 579–586. [CrossRef]
4. Graf, R. New Possibilities for the Diagnosis of Congenital Hip Joint Dislocation by Ultrasonography. *J. Pediatric Orthop.* **1983**, *3*, 354–359. [CrossRef]
5. Shorter, D.; Hong, T.; Osborn, D.A. Cochrane Review: Screening Programmes for Developmental Dysplasia of the Hip in Newborn Infants. *Evid.-Based Child Health* **2013**, *8*, 11–54. [CrossRef]
6. Sink, E.L.; Ricciardi, B.F.; Dela Torre, K.; Price, C.T. Selective Ultrasound Screening Is Inadequate to Identify Patients Who Present with Symptomatic Adult Acetabular Dysplasia. *J. Child. Orthop.* **2014**, *8*, 451–455. [CrossRef]
7. Choudry, Q.A.; Paton, R.W. Neonatal Screening and Selective Sonographic Imaging in the Diagnosis of Developmental Dysplasia of the Hip. *Bone Jt. J.* **2018**, *100*, 806–810. [CrossRef]
8. Husum, H.C.; Thomsen, J.L.; Kold, S.; Maimburg, R.D.; Rahbek, O. Referral Criteria Recognition of Screeners in the Danish Screening Programme for Hip Dysplasia. *Dan. Med. J.* **2022**, *69*, A01210098.
9. Husum, H.C.; Maimburg, R.D.; Kold, S.; Thomsen, J.L.; Rahbek, O. Self-Reported Knowledge of National Guidelines for Clinical Screening for Hip Dysplasia: A Web-Based Survey of Midwives and GPs in Denmark. *BJGP Open* **2021**, *5*, 1–8. [CrossRef]

10. Broadhurst, C.; Rhodes, A.M.L.; Harper, P.; Perry, D.C.; Clarke, N.M.P.; Aarvold, A. What Is the Incidence of Late Detection of Developmental Dysplasia of the Hip in England?: A 26-Year National Study of Children Diagnosed after the Age of One. *Bone Jt.*
J. **2019**, *101*, 281–287. [CrossRef]
11. Thaler, M.; Biedermann, R.; Lair, J.; Krismer, M.; Landauer, F. Cost-Effectiveness of Universal Ultrasound Screening Compared with Clinical Examination Alone in the Diagnosis and Treatment of Neonatal Hip Dysplasia in Austria. *J. Bone Jt. Surg.-Ser. B* **2011**, *93*, 1126–1130. [CrossRef] [PubMed]
12. O'Beirne, J.G.; Chlapoutakis, K.; Alshryda, S.; Aydingoz, U.; Baumann, T.; Casini, C.; de Pellegrin, M.; Domos, G.; Dubs, B.; Hemmadi, S.; et al. International Interdisciplinary Consensus Meeting on the Evaluation of Developmental Dysplasia of the Hip Internationales Interdisziplinares Konsensustreffen Zur Evaluation Der Diagnostik Und Therapie Der Angeborenen Hüftdysplasie Authors. *Ultraschall Med.* **2019**, *40*, 454–464. [PubMed]
13. Tréguier, C.; Chapuis, M.; Branger, B.; Bruneau, B.; Grellier, A.; Chouklati, K.; Proisy, M.; Darnault, P.; Violas, P.; Pladys, P.; et al. Pubo-Femoral Distance: An Easy Sonographic Screening Test to Avoid Late Diagnosis of Developmental Dysplasia of the Hip. *Eur. Radiol.* **2013**, *23*, 836–844. [CrossRef]
14. Husum, H.-C.; Hellfritzsch, M.B.; Hardgrib, N.; Møller-Madsen, B.; Rahbek, O. Suggestion for New 4.4 Mm Pubo-Femoral
Distance Cut-off Value for Hip Instability in Lateral Position during DDH Screening. *Acta Orthop.* **2019**, *3674*, 88–93. [CrossRef]
15. von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *J. Clin. Epidemiol.* **2008**, *61*, 344–349. [CrossRef] [PubMed]
16. Miller, G.E. The Assessment of Clinical Skills/Competence/Performance. *Acad. Med.* **1990**, *65*, S63–S67. [CrossRef]
17. Salut, C.; Moriau, D.; Pascaud, E.; Layré, B.; Peyrou, P.; Maubon, A. Initial Results from an Ultrasound Screening Program for the Detection of Developmental Dysplasia of the Hip in Girls. *J. Radiol.* **2011**, *92*, 920–929. [CrossRef]
18. Portney, L.G.; Watkins, M.P. Foundation Clinical Research. *Foreign Aff.* **2012**, *91*, 1689–1699. [CrossRef]
19. Costa-Santos, C.; Bernardes, J.; Ayres-De-Campos, D.; Costa, A.; Costa, C. The Limits of Agreement and the Intraclass Correlation Coefficient May Be Inconsistent in the Interpretation of Agreement. *J. Clin. Epidemiol.* **2011**, *64*, 264–269. [CrossRef]
20. Teixeira, S.R.; Dalto, V.F.; Maranho, D.A.; Zoghbi-Neto, O.S.; Volpon, J.B.; Nogueira-Barbosa, M.H. Comparison between Graf Method and Pubo-Femoral Distance in Neutral and Flexion Positions to Diagnose Developmental Dysplasia of the Hip. *Eur. J. Radiol.* **2015**, *84*, 301–306. [CrossRef]
21. Quader, N.; Schaeffer, E.K.; Hodgson, A.J.; Abugharbieh, R.; Mulpuri, K. A Systematic Review and Meta-Analysis on the Reproducibility of Ultrasound-Based Metrics for Assessing Developmental Dysplasia of the Hip. *J. Pediatr. Orthop.* **2018**, *38*, e305–e311. [CrossRef] [PubMed]

22. Jaremko, J.L.; Mabee, M.; Swami, V.G.; Jamieson, L.; Chow, K.; Thompson, R.B. Potential for Change in US Diagnosis of Hip Dysplasia Solely Caused by Changes in Probe Orientation: Patterns of Alpha-Angle Variation Revealed by Using ThreeDimensional US. *Radiology* **2014**, *273*, 870–878. [CrossRef] [PubMed]
23. Orak, M.M.; Onay, T.; Cagirmaz, T.; Elibol, C.; Elibol, F.D.; Centel, T. The Reliability of Ultrasonography in Developmental
Dysplasia of the Hip: How Reliable Is It in Different Hands? *Indian J. Orthop.* **2015**, *49*, 610–614. [CrossRef] [PubMed]

5.2. APPENDIX II

Point of care pubo-femoral distance ultrasound outperforms traditional selective screening for hip dysplasia

Hans-Christen Husum[a,b], MD, Michel Bach Hellfritzsch[b,c], MD, Rikke Damkjær Maimburg[b,e], MPH, PhD, Bjarne Møller-Madsen[b,d], MD, DMSC Mads Henriksen[b,c], MD, Natallia Lapitskaya[b,c], MD, PhD, Søren Kold[a,b] MD, PhD, Ole Rahbek[a,b], MD, PhD.

Affiliations

[a]Interdisciplinary Orthopaedics, Aalborg University Hospital, Denmark

[b]Danish Paedatric Orthopaedic Research, Aarhus University Hospital, Denmark www.dpor.dk

[c]Department of Radiology, Aarhus University Hospital, Denmark

[d]Department of Children's Orthopaedics, Aarhus University Hospital, Denmark

[e]Department of Midwifery, University College of Northern Denmark, Aalborg

Corresponding author: Hans-Christen Husum, Interdisciplinary Orthopaedics, Aalborg University Hospital, Hobrovej 18-22, 9000 Aalborg, Denmark. Email: husumcorr@hotmail.com

Conflict of interest: The authors have no conflicts of interest relevant to this article to disclose.

Funding/support: This study was funded by: The Danish Arthritis Foundation, the Dagmar Marshall Fund and the Independent Research Fund Denmark Grant no. 1030-00366B

Author contributions

Conceptualization: all authors

Methodology: Drs Husum and Rahbek

Formal analysis: Dr Husum

Investigation: Drs Husum, Hellfritzsch, Henriksen and Lapitskaya

Data curation: Dr Husum

Writing – original draft: Drs Husum and Rahbek

Writing – review & editing: All authors

Supervision: Dr Rahbek and Ms Maimburg

Project administration: Drs Husum, Rahbek, Møller-Madsen and Ms Maimburg.

Funding acquisition: Drs Husum and Rahbek.

Abstract

Aims

To investigate and compare the effectiveness of primary PFD screening as a referral criterion for follow-up (FU) Graf hip ultrasound (US), to traditional referral criteria in selective US screening for developmental dysplasia of the hip (DDH). Secondly, to propose an optimal cut-off value for primary PFD screening in the lateral examination position.

Methods

We prospectively included all newborns consented to receive a primary PFD US screening of both hips in the selective screening programme at our institution. The PFD criterion was compared to traditional referral criteria in terms of sensitivity, specificity and positive predictive values (PPV) in detecting abnormal hips on FU US, as well as proportion of abnormal hips found and referral rates.

Results

We included 2,735 newborns of which 616 received a FU hip ultrasound. After exclusion 561 newborns were included for analysis. Gender distribution was 283 female and 273 male, mean age at FU US was 36.6 days (range 4-87 days). 317 newborns (11.6%) were referred by traditional screening criteria and 303 newborns (10.8%) were referred by the PFD criterion.

Sensitivities/specificities for detecting ≥Graf type IIa hips were: 17.4%/94.2% for clinical examination, 27.9%/47.5% for risk factors, 40.7%/51% for clinical examination and risk factors combined and 65.1%/72% for PFD examination using a cut-off of 5.8 mm. Differences in sensitivities and specificities between traditional referral criteria and the PFD criterion were statistically significant ($p<0.01$).

Conclusion

Early point-of-care PFD US screening has significantly higher sensitivity, specificity and PPV in detecting abnormal hips, when compared to current selective criteria. PFD US increases the detection rate of immature hips (Graf IIa) by 72% and dysplastic hips (≥ Graf IIc) by 60% while maintaining similar referral rates

Introduction

Developmental dysplasia of the hip (DDH) is a condition defined by underdevelopment of the hip joint which can lead to pain, hip joint instability and early osteoarthritis [1]. With an incidence of 0.8% [2] of all newborn children it is the most common musculoskeletal developmental disorder in children.

Universal ultrasound (US) screening of all children is increasingly favoured over selective US screening [3], but is difficult and time-consuming to implement on a large scale, due to financial and logistical considerations [4]. Hip US is commonly based on the methodology developed by Graf [5], but the complexity of the Graf method and its requirements, necessitates high levels of training of examiners with subsequent monitoring to ensure the diagnostic quality of examiners using the method [6], which may cause a strain on specialised health care professionals and delay the implementation process.

A sensitive, reproducible and easily taught point of care (POC) US method could therefore ease the implementation of a primary US screening programme and provide a feasible alternative to primary screening based on the more complex Graf method.

In 2013 the Pubo-femoral distance (PFD) was suggested as an US screening measurement [7]. The PFD method is highly sensitive for DDH in the hands of skilled examiners [8] and has high reliability even when performed by novice users [9]. To date no studies have evaluated the PFD method as a referral criterion for follow-up (FU) Graf hip US and compared it to traditional selective referral criteria in DDH screening.

The aim of the present study is to evaluate the performance of PFD measurements as a referral criterion for FU Graf hip US and to compare it to the performance of clinical hip examinations and risk factor identification currently used as referral criteria in selective screening programmes worldwide. We will compare the PFD criterion to traditional selective referral criteria in terms of sensitivity, specificity, and positive predictive values (PPV) in detecting US positive DDH, the proportion of abnormal hips found and referral rates. Secondly, we propose an optimized cut-off value for the detection of DDH in early PFD US examination of newborn hips.

Materials and Methods

Setting

Design: Prospective interventional study running from October 2021 to October 2022 at Aarhus University Hospital (AUH), Denmark.

The referral criteria for FU hip US in the screening programme for DDH at AUH is based on clinical hip examinations based on the Ortolani and Barlow manoeuvres and risk factor identification (family history, breech presentation, oligohydramnios, clubfeet, musculoskeletal syndromes). All newborns born at AUH receive a clinical- and risk factor screening performed by a midwife in the post-partum clinic when the child is approximately two days old.

Parents of newborns undergoing screening were offered a hip US with PFD measurement of their newborns' hips as a supplement to the primary clinical examination and risk factor identification. The PFD US was performed in a separate room, ideally on the same day as the clinical examination, by another midwife trained in PFD measurements. PFD examination was performed using a MINDRAY TE7 US scanner and a high frequency (16MHz) linear transducer (Mindray Medical International, Shenzhen, China). PFD was measured according to the methods described by Treguiér and Couture, but with the child in the lateral examination position [710] (Figure 1). A description of the training programme, the PFD method and reliability of midwife-performed PFD measurements has been documented in a previous publication [9].

PFD screening was performed on the same day as clinical screening or, in the case of clinical screening being performed in the weekend, in the following week. To negate any bias in comparison between screening criteria introduced by the interval between clinical- and PFD examination, the PFD screening had to be completed within 14 days after birth.

The combined referral criteria for a FU hip US were: a positive clinical finding upon hip examination and/or a risk factor for DDH and/or PFD above the 5.1mm threshold or a difference in PFD between both hips of 1.5mm or above (figure 2). FU hip US was performed by one of three musculoskeletal radiologists according to the methods by Graf and Harcke [511] (figure 1). FU hip US would take place ideally before the child was six weeks of age, or before two weeks of age if clinical instability was detected upon examination or, per study protocol, if the PFD measurement was above 8 mm, indicating US instability.

We estimated an initial PFD cut-off for detecting abnormal hips (≥Graf type IIa) based upon retrospective analysis of PFD measurements performed at AUH and the existing literature [7128]. We decided on a low cut-off at 5.1mm to increase the proportion of

newborns referred and thus the amount of data for analysis. Further, we included the original authors suggestion of a PFD difference of 1.5mm between both hips as a referral criterion [7]. An updated optimal PFD cut-off was determined post-hoc and the results reported in this study uses the updated post-hoc cut-off value as well as the 1.5mm PFD difference.

Participants

Inclusion criteria: newborns < 14 days old screened at the post-partum clinic at AUH with written parental consent for participation.

We excluded newborns who were referred to other institutions for their FU hip US.

Study size

Sample size calculation was based on differences in Graf type IIc hip detection rates between a selective screening programme at AUH and published results from a universal screening programme in Austria [13]. A difference in Graf IIc detection rates of 0.46% was found and with a power of 90 and alfa of 5% a sample size of 2.204 newborns was calculated.

Statistical methods

To compare referral criteria, a ROC curve for each referral criteria (clinical examination, risk factor and PFD) was created while confidence intervals sensitivity, specificity and AUC values were calculated using 100 bootstrap samples. Positive predictive values for each criterion were calculated as true positives divided by all referrals with a true positive defined as ≥ Graf type IIa hips. Sensitivity and specificity was compared using the Mcnemar test and correlated AUC values were calculated according to the method by Delong [14]. Further, referral rates and results from the FU hip US stratified by Graf hip classifications were calculated for the current selective screening criteria (clinical examination and/or risk factor) and the PFD criterion. To account for bilaterality in data, a sensitivity analysis was performed, by calculating empirical cut-off values for left, right and all hips. As no significant difference was detected in the sensitivity analysis, independency between sides was assumed. Normal distribution of data was investigated using QQ plots and a significance level of 5% was used.

Determining optimal PFD cut-off for analysis

We estimated an optimal cut-off value for the PFD post-hoc by calculating empirical cut-off points of 5.8 and 6.1 mm using the Youden and Liu indexes [1516]. To calculate prediction errors and confidence intervals of the cut-off values we performed a repeated 10-fold cross validation while assuming independency between hips. A final

value of 5.8mm was chosen by the first- and senior author as the optimal cut-off point upon inspection of cut-off values and their corresponding sensitivities and specificities (supplementary table 1). The reported results in this study were calculated using the 5.8mm PFD threshold while keeping the 1.5 mm difference in PFD between hips.

Ethics

The study and the written patient information leaflet for the parents were approved by the local institutional ethics committee (Ref no: N-20200051).

Funding

The project was funded by The Independent Research Fund Denmark, The Danish Arthritis foundation, The Health professional fund at AUH and the Dagmar Marshall Fund Grant no. 1030-00366B.

Results

During the study period 4,794 newborns were screened at the post-partum clinic. We included 2,735 newborns, 2,094 newborns were not referred for specialised US as they did not fulfil criteria of the selective screening programme and had a PFD below 5.8 mm. 25 newborns were above 14 days of age at PFD screening, 47 referred newborns did not show up for FU US, and eight newborns were referred to other institutions in the region leaving 561 children for analysis (figure 3).

Gender distribution was 283 female and 273 males, mean age at primary PFD screening was 3.7 days (range 0-13 days), and mean age at FU hip US was 35.6 days (range 4-121 days). 317 newborns were referred by traditional screening criteria and 303 newborns were referred by the PFD criterion (≥5.8 mm) corresponding to referral rates of 11.6% and 10.8% respectively.

Sensitivities for detecting ≥Graf type IIa hips were: 17.4% (95% CI 9.9% ; 24.0%) for clinical examination, 27.9% (95% CI 18.8% ; 39.4%) for risk factors, 40.7% (95% CI 28.9% ; 50.4%) for clinical examination and risk factors combined and 65.1% (95% CI 58.5% ; 76.1%) for PFD examination using a cut-off of 5.8 mm. Specificities were 94.2% (95% CI 52% ; 59%) for clinical examination, 47.5% (95% CI 43% ; 54%) for risk factors, 51% (95% CI 45% ; 56%) for clinical examination and risk factors combined and 72% (95% CI 68%;77%) for PFD examination.

Differences in sensitivities and specificities between traditional referral criteria and the PFD criterion were statistically significant ($p<0.01$). A detailed comparison including PPV and AUC values for all referral criteria are presented in table 1 and figure 5.

Classification of hips referred by combined clinical examination and risk factor identification were: 548 Type I, 29 type IIa, 5 type IIc and 1 ≥ type III, for PFD examination the corresponding findings were: 420 type I, 50 type IIa, 8 type IIc and 1 ≥ type III. A detailed presentation of hip classifications and referral rates stratified by screening criteria can be seen in figure 4.

Discussion

This study found a significantly higher sensitivity, specificity and PPV for detecting abnormal hips, when comparing PFD US examination to traditional screening criteria in a population of 2,730 newborns. PFD examination detected an additional 21 type IIa hips and three type IIc hips, translating to increases of 72% and 60%, respectively with no increase in referral rates. A PFD cut-off value of 5.8mm was determined as the optimal cut-off value in early PFD US screening of newborns for DDH.

Interpretation

Timing of US screening for DDH has been found to be unreliable before the fourth week of life [17], while others have found that US screening within three days of life reliably detected DDH[18]. In the present study, while early PFD US screening was performed within the first days of life, the final diagnosis of DDH was made at the six-week FU US, which follows international recommendations [19].

We chose to define immature Graf type IIa hips as abnormal, and consequently designed our study to detect these hips. While 95% of immature hips spontaneously resolve with time, 5% do not [20] and it is therefore recommended to do FU on all immature hips until maturation [21]. Until prediction of which hips will resolve on their own is possible, a DDH screening programme should therefore be able to detect hip immaturity, even though it will have no consequence in treatment for most patients.

Cut-off values for PFD in DDH screening have been previously proposed, but the definition of what constitutes true DDH varies and the application of the US method is thus not consistent. [22] [78]. The original authors suggested a cut-off of 6.0 mm or a difference between hips of 1.5mm measured with the child in the supine position. While they did not describe in detail how they arrived at this cut-off, implementing DDH screening, using this cut-off, reportedly reduced late DDH diagnoses to 0%. Motta and colleagues suggested a cut-off of 3.0mm measured in the lateral position, but did not submit the hip to the lateral stress described in the method by Couture [23]. We previously suggested a cut-off of 4.4mm when detecting dysplastic hips in need of treatment in older children. The presently suggested 5.8mm cut-off value is therefore the first to be proposed for use in early US screening for DDH. As any decision of cut-off values is a compromise of gained sensitivity versus lost specificity, we have added a table of cut-off values and corresponding sensitivities and specificities (supplementary table 1), as others may disagree with our choice.

For any screening program to be feasible, the screening method needs to be reliable, implementable and have an acceptable rate of detection, while not producing an unacceptable number of false positives. While selective screening based on clinical hip examinations and risk factor screening is already widely implemented, the predictive value and knowledge of clinical hip examinations, in a general pool of

screeners, is low [24] [25] which increases referral rates without a similar increase in treatment [26]. Further, 51% of newborns with DDH do not have positive physical signs or risk factors for DDH[18]. Similarly, in the present study 53% of abnormal hips were detected in newborns who had negative physical findings and no risk factors for DDH.

Universal US screening using the Graf method has been proposed as an alternative to clinical screening, but the Graf method has a low reliability [27] and requires extensive training and quality monitoring to be performed correctly [6], challenging large scale implementation. Conversely, the PFD method may be the useful compromise, as it is reliable [7] [28], and swiftly be taught to even novice US users [28] [9]. Although the detection rates may be lower than those seen in universal Graf US screening programmes, the sensitivities and specificities are superior to those of clinical- and risk factor screening as presently demonstrated.

Limitations

Participation in the study provided the opportunity for parents to receive a referral for a specialized hip US if the PFD examination was positive. As such, we observed that parents of children who were already referred, based on a positive clinical examination or presence of risk factors, were likely to not participate in the study. Further, as the PFD screening took place in a separate room from the post-partum clinic, we anecdotally experienced that parents were reluctant to relocate their newborn to another room for secondary hip screening procedure. While this behaviour was not systematically registered, and we do not have any information on families that did not participate, it may cause a selection bias in our study population which underrepresents previously referred children, and complicates the comparison of our findings to those of other screening programmes.

External validity

As all newborns did not receive a FU hip US, per the selective screening guidelines at AUH, the reported sensitivities and specificities should not be considered valid outside this study population. As we have no information on how many abnormal hips were missed in our cohort, the values are only presented to enable a comparison between screening criteria.

Conclusion

Early PFD US screening has significantly higher sensitivity, specificity and PPV in detecting abnormal hips in newborns compared to traditional selective criteria. PFD US increases the detection rate of immature hips by 72% and dysplastic hips by 60% with similar referral rates. Early PFD screening as a POC method may be considered an alternative to selective screening programmes for DDH.

References

1. **Engesæter IØ, Lie SA, Lehmann TG, Furnes O, Vollset SE, Engesæter LB, et al.** Neonatal hip instability and risk of total hip replacement in young adulthood: Follow-up of 2,218,596 newborns from the Medical Birth Registry of Norway in the Norwegian Arthroplasty Register. *Acta Orthop* 2008;79(3):321–326.

2. **Biedermann R, Riccabona J, Giesinger JM, Brunner A, Liebensteiner M, Wansch J, et al.** Results of universal ultrasound screening for developmental dysplasia of the hip. *Bone Joint J* [Internet] 2018 [cited 17 Jan 2020];100-B(10):1399–1404.

3. **O'Beirne JG, Chlapoutakis K, Alshryda S, Aydingoz U, Baumann T, Casini C, et al.** International Interdisciplinary Consensus Meeting on the Evaluation of Developmental Dysplasia of the Hip Internationales interdisziplinares Konsensustreffen zur Evaluation der Diagnostik und Therapie der angeborenen Hüftdysplasie Authors. *Ultraschall Med* [Internet] 2019;40(4):454–464.

4. **Thaler M, Biedermann R, Lair J, Krismer M, Landauer F, Landauer F**. Cost-effectiveness of universal ultrasound screening compared with clinical examination alone in the diagnosis and treatment of neonatal hip dysplasia in Austria. *J Bone Jt Surg - Ser B* 2011;93 B(8):1126–1130.

5. **Graf R**. New Possibilities for the Diagnosis of Congenital Hip Joint Dislocation by Ultrasonography. Journal of Pediatric Orthopedics, 1983.

6. **Ömeroğlu H**. Significant Assessment Discrepancies Occur Between the Examiners Having and Not Having Further Special Hands-on Course Trainings in Infantile Hip Ultrasonography by the Graf Method. *Indian J Orthop* [Internet] 2021;

7. **Tréguier C, Chapuis M, Branger B, Bruneau B, Grellier A, Chouklati K, et al.** Pubo-femoral distance: An easy sonographic screening test to avoid late diagnosis of developmental dysplasia of the hip. *Eur Radiol* 2013;23(3):836–844.

8. **Husum H-C, Hellfritzsch MBMB, Hardgrib N, Møller-Madsen B, Rahbek O**. Suggestion for new 4.4 mm pubo-femoral distance cut-off value for hip instability in lateral position during DDH screening. *Acta Orthop* [Internet] 2019;3674(1):88–93.

9. **Husum H-C, Bach Hellfritzsch M, Maimburg RD, Henriksen M, Lapitskaya N,**

Møller-Madsen B, et al. Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound. *Children* [Internet] 2022;9(9):1345.

10. **Couture A, Baud C, Prodhomme O, Saquintaah M, Veyrac C**. Ultrasound of the neonatal hip: initial evaluation and follow-up. *J Radiol* [Internet] 2011;92(2):142–165.

11. **Harcke HT, Clarke NM, Lee MS, Borns PF, MacEwen GD**. Examination of the infant hip with real-time ultrasonography. *J Ultrasound Med* [Internet] 1984;3(3):131–7.

12. **Ban Y, Luan Q, Shi M, Sun B, Li T, Wang Y, et al.** Establishing reference values for the pubofemoral distance in normal infant medial hips by ultrasound. *Acta radiol* 2020;(324).

13. **Biedermann R, Eastwood DM**. Universal or selective ultrasound screening for developmental dysplasia of the hip? A discussion of the key issues. *J Child Orthop* 2018;12(4):296–301.

14. **Delong ER, Delong DM, Clarke-Pearson DL**. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;837–45.

15. **Smits N**. (Smits 2010) A note on Youden's J and its cost ratio. 2010;2–5.

16. **Liu X**. Classification accuracy and cut pointselection. *Stat Med* 2012;31(23):2676–2686.

17. **Tan SHS, Wong KL, Lim AKS, Hui JH**. The earliest timing of ultrasound in screening for developmental dysplasia of the hips. *Ultrasonography* 2019;38(4):321–326.

18. **Gyurkovits Z, Sohár G, Baricsa A, Németh G, Orvos H, Dubs B**. Early detection of developmental dysplasia of hip by ultrasound. *HIP Int* [Internet] 2019;112070001987968.

19. **Mulpuri K, Song KM, Gross RH enr., Tebor GB, Otsuka NY oshinob., Lubicky JP, et al.** The American Academy of Orthopaedic Surgeons Evidence-Based Guideline on Detection and Nonoperative Management of Pediatric Developmental Dysplasia of the Hip in Infants up to Six Months of Age. *J Bone Joint Surg Am* 2015;97(20):1717–1718.

20. **Roovers EA, Boere-Boonekamp MM, Mostert AK, Castelein RM, Zielhuis GA, Kerkhoff THM**. The natural history of developmental dysplasia of the

hip: Sonographic findings in infants of 1-3 months of age. *J Pediatr Orthop Part B* 2005;14(5):325–330.

21. **Graf R**. *Hip Sonography. Diagnosis and Management of Infant Hip Dysplasia*. 2nd ed. Springer. Springer, 2006.

22. **Motta GGB, Rodrigues A, Chiovatto S, Chiovatto ED, Takahashi MS, Vogel N, et al.** Measurement of Pubofemoral Distance in the Diagnosis of Developmental Dysplasia of the Hip.

23. **Salut C, Moriau D, Pascaud E, Layré B, Peyrou P, Maubon a**. [Initial results from an ultrasound screening program for the detection of developmental dysplasia of the hip in girls]. *J Radiol* [Internet] 2011;92(10):920–9.

24. **Choudry QA, Paton RW**. Neonatal screening and selective sonographic imaging in the diagnosis of developmental dysplasia of the hip. *Bone Jt J* 2018;100B(6):806–810.

25. **Husum HC, Maimburg RD, Kold S, Thomsen JL, Rahbek O**. Self-reported knowledge of national guidelines for clinical screening for hip dysplasia: a web-based survey of midwives and GPs in Denmark. *BJGP Open* 2021;5(4):1–8.

26. **Düppe H, Danielsson LG**. Screening of neonatal instability and of developmental dislocation of the hip. A survey of 132,601 living newborn infants between 1956 and 1999. *J Bone Joint Surg Br* [Internet] 2002;84(6):878–85.

27. **Quader N, Schaeffer EK, Hodgson AJ, Abugharbieh R, Mulpuri K**. A Systematic Review and Meta-analysis on the Reproducibility of Ultrasound-based Metrics for Assessing Developmental Dysplasia of the Hip. *J Pediatr Orthop* [Internet] United States, 2018;38(6):e305–e311.

28. **Teixeira SR, Dalto VF, Maranho DA, Zoghbi-Neto OS, Volpon JB, Nogueira-Barbosa MH**. Comparison between Graf method and pubo-femoral distance in neutral and flexion positions to diagnose developmental dysplasia of the hip. *Eur J Radiol* [Internet] Elsevier Ireland Ltd, 2015;84(2):301–306.

Figures and tables

Table 1: sensitivities, specificities, AUC values and PPVs stratified by referral criteria for DDH screening. P values compare estimates for previous row of referral criteria. PFD = pubo-femoral distance, AUC = Area under the curve, PPV = Positive predictive value.

Referral criteria	**Sensitivity (95% CI)**	**P-value**	**Specificity % (95% CI)***	**P-value**	**AUC (95% CI)**	**P-value**	**PPV % (95% CI)**
Clinical examination	17.4% (9.9% ; 24.0%)	-	94.2% (93.0% ; 95.4%)	-	56% (52% ; 59%)	-	17.4% (9.9% ; 24.0%)
Risk factor	27.9% (18.8% ; 39.4%)	0.15	67.1% (65.0% ; 69.1%)	<0.001	47.5% (43% ; 54%)	0.011	4.9% (3.0% ; 6.9%)
Clinical exam and/or risk factor	40.7% (28.9% ; 50.4%)	<0.001	61.7% (59.0% ; 63.9%)	<0.001	51% (45% ; 56%)	0.049	6.0% (4.1% ; 8.3%)
PFD (≥5.8mm)	65.1% (58.5% ; 76.1%)	0.004	78.0% (75.6% ; 80.1%)	<0.001	72% (68% ; 77%)	<0.001	15.1% (12.4% ; 18.8%)

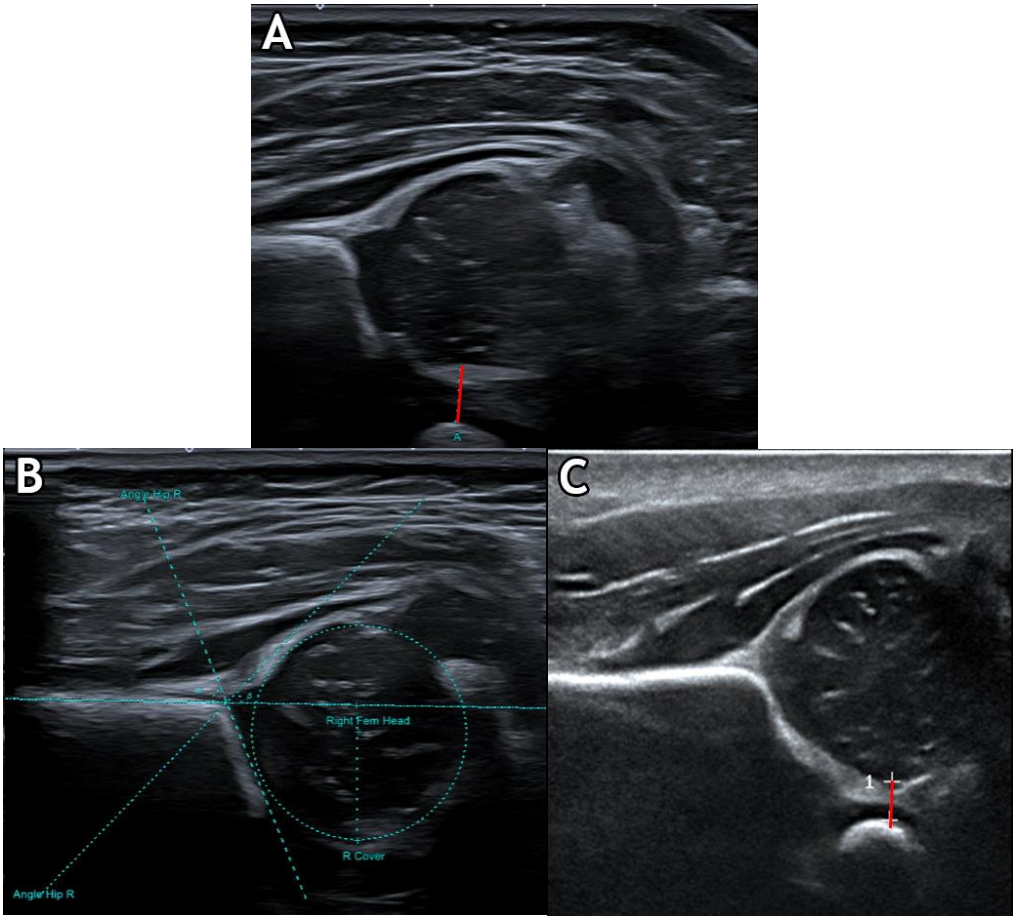


Figure 1: Ultrasound images of three newborn hips as performed by a radiologist (A,B) and midwife (C). (A) and (C) depicts the pubo-femoral distance (PFD) method with PFD marked as a red line, while (B) depicts the Graf method with annotated alpha angles.

Figure 2: Flowchart depicting the screening process of the newborns who were consented to participate in the study. PFD = pubo femoral distance.

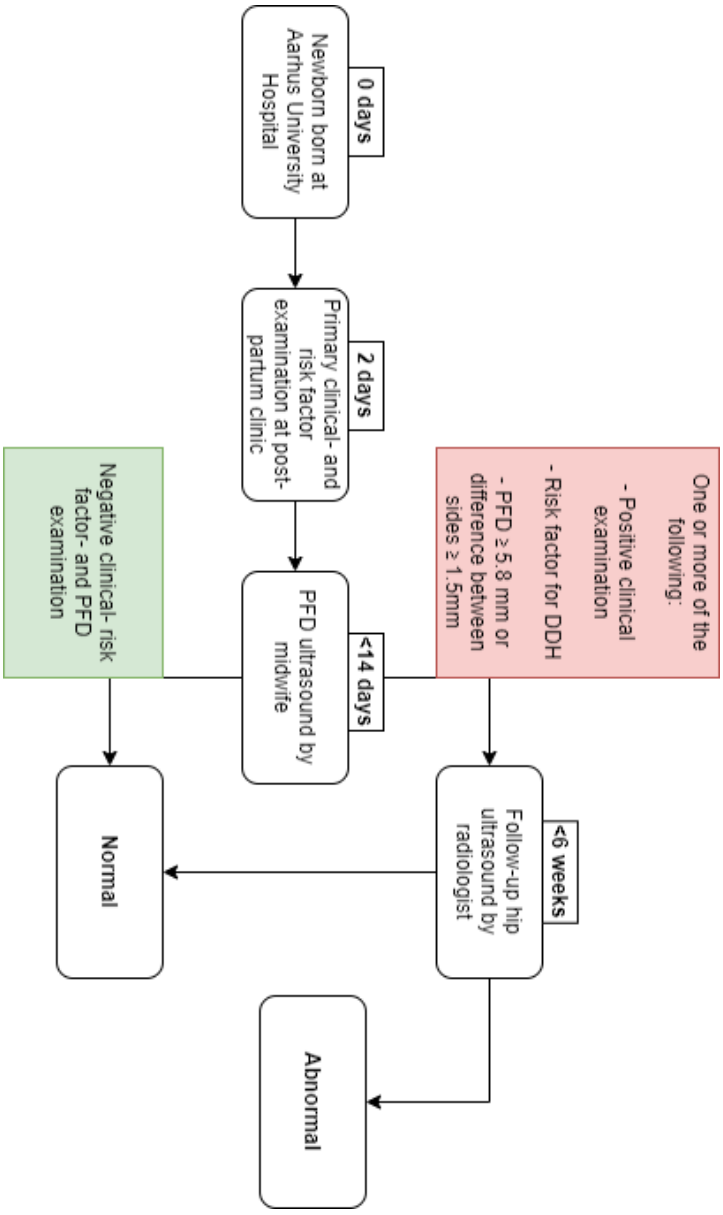


Figure 3: CONSORT diagram of the inclusion process. PFD = pubo-femoral distance

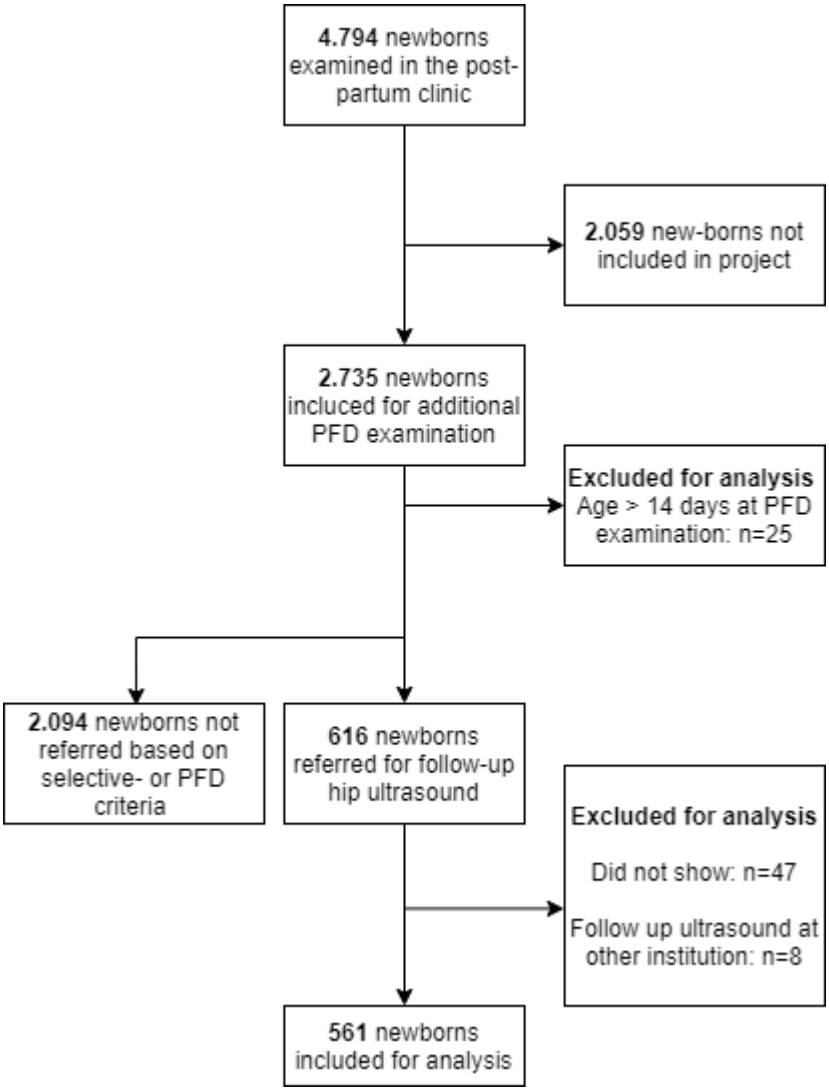


Figure 4: Ultrasound findings for each set of screening criteria for the current study. The green and blue boxes the total number of patients and findings in each program. Referral numbers for the PFD program reflect a cut off value of PFD ≥ 5.8mm or a difference in PFD ≥1.5mm between hips.

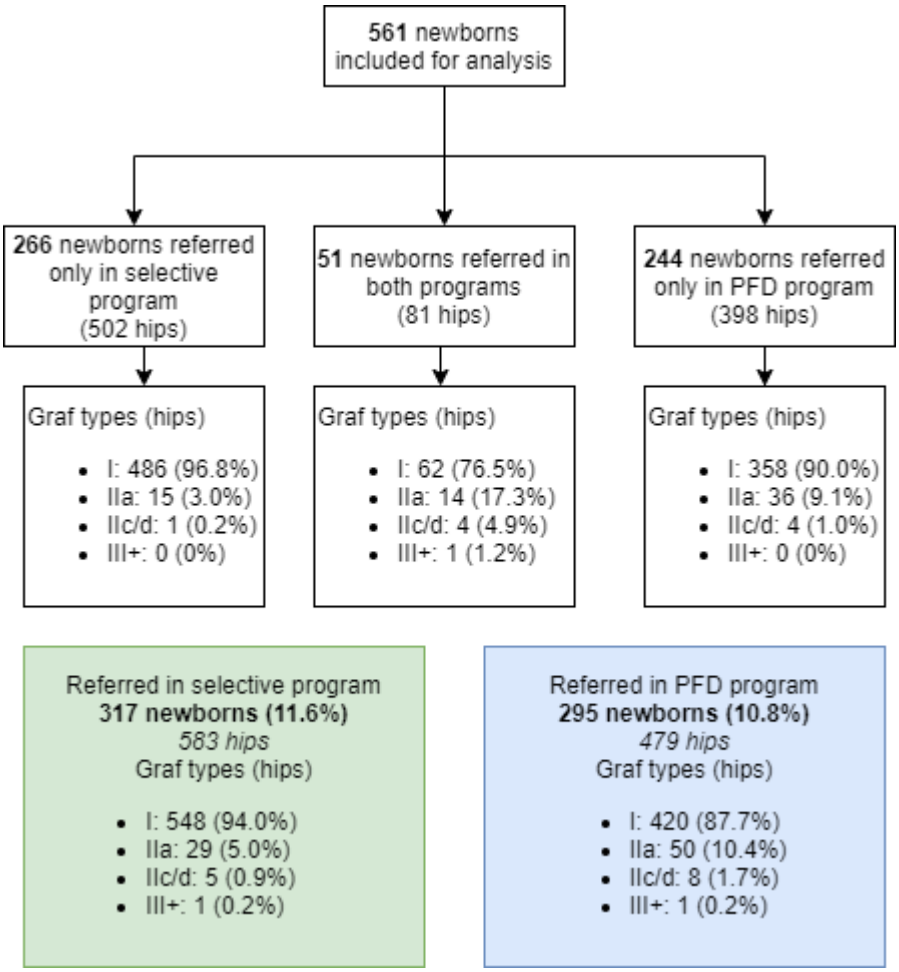
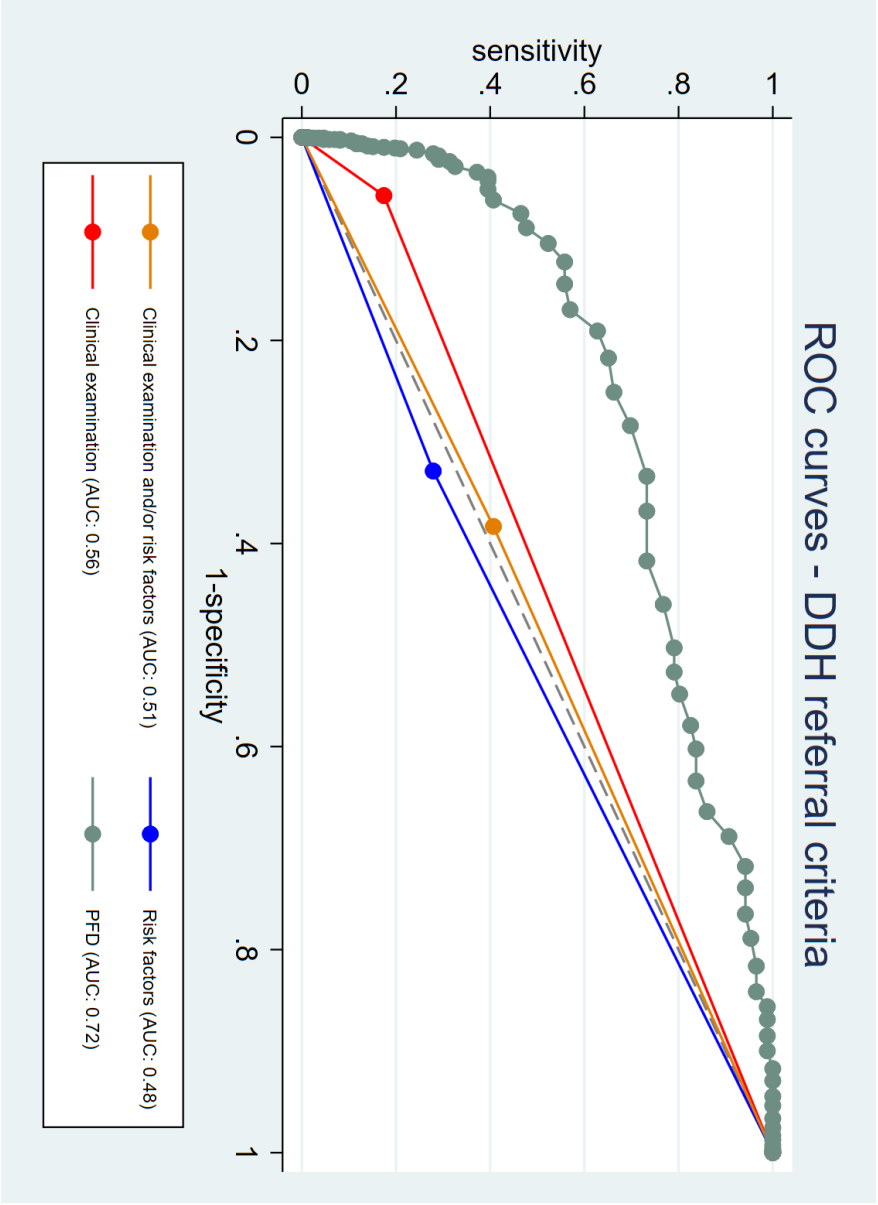


Figure 5: Receiver operating characteristics curves depicting sensitivities and 1-specifities for each referral criteria analysed. The single point for the dichotomous referral criteria (orange, blue and red) represents the sensitivity and 1-specificty when the referral criteria is obtained. The multiple points for the PFD measurement (teal) represents the sensitivity and 1-specificity for each cut-off value of the primary PFD measurement. PFD = pubo-femoral distance, AUC = Area under the curve, DDH = Developmental dysplasia of the hip.



5.3. APPENDIX III

Pubo femoral distances correlate to acetabular morphology and sonographic instability in screening for hip dysplasia

Hans-Christen Husum[a,b], MD, Michel Bach Hellfritzsch[b,c], MD, Rikke Damkjær Maimburg[b,e], MPH, PhD, Bjarne Møller-Madsen[b,d], MD, DMSC Mads Henriksen[b,c], MD, Natallia Lapitskaya[b,c], MD, PhD, Søren Kold[a,b] MD, PhD, Ole Rahbek[a,b], MD, PhD.

Affiliations

[a]Interdisciplinary Orthopaedics, Aalborg University Hospital, Denmark

[b]Danish Paedatric Orthopaedic Research, Aarhus University Hospital, Denmark www.dpor.dk

[c]Department of Radiology, Aarhus University Hospital, Denmark

[d]Department of Children's Orthopaedics, Aarhus University Hospital, Denmark

[e]Department of Midwifery, University College of Northern Denmark, Aalborg

Corresponding author: Hans-Christen Husum, Interdisciplinary Orthopaedics, Aalborg University Hospital, Hobrovej 18-22, 9000 Aalborg, Denmark. Email: husumcorr@hotmail.com

Conflict of interest: The authors have no conflicts of interest relevant to this article to disclose.

Funding/support: This study was funded by: The Danish Arthritis Foundation, the Dagmar Marshall Fund and the Independent Research Fund Denmark Grant no. 1030-00366B

Author contributions

Conceptualization: all authors

Methodology: Drs Husum and Rahbek

Formal analysis: Dr Husum

Investigation: Drs Husum, Hellfritzsch, Henriksen and Lapitskaya

Data curation: Dr Husum

Writing – original draft: Drs Husum and Rahbek

Writing – review & editing: All authors

Supervision: Dr Rahbek and Ms Maimburg

Project administration: Drs Husum, Rahbek, Møller-Madsen and Ms Maimburg.

Funding acquisition: Drs Husum and Rahbek.

Abstract

Aims

The present study seeks to investigate the correlation of pubo-femoral distances (PFD) to alpha angles, and hip displaceability status defined as femoral head coverage (FHC) or FHC during manual provocation of the newborn hip < 50%.

Methods

We retrospectively included all newborns referred for ultrasound screening at our institution based on primary risk factor-, clinical- and PFD screening. Alpha angles, PFD, FHC and FHC at follow-up ultrasound for referred newborns were measured and compared using scatter plots, linear regression, t-tests and box-plots.

Results

We included 2,735 newborns of which 754 received a follow-up hip ultrasound within six weeks of age. After exclusion 1,500 hips were included for analysis. Gender distribution was 372 male 380 female, mean age at examination was 36.6 days (range 4-87 days).

We found a negative linear correlation of PFD to alpha angles ($p<0.001$), FHC($p<0.001$) and FHC during provocation ($p<0.001$) with a 1mm increase in PFD corresponding to a -2.1 degree (95% CI -2.3;-1.9) change in alpha angle and a -3.4% (95% CI -3.7;-3.0) change in FHC and a -6.0% (-6.6;-5.5) change in FHC during provocation. The PFD was significantly higher with increasing Graf types and in displaceable hips ($p<0.001$)

Conclusion

PFD is strongly correlated to both alpha angles and hip displaceability, as measured by FHC and FHC during provocation, in ultrasound of newborn hips. The PFD increases as the hips become more dysplastic and/or displaceable.

Clinical relevance

- The PFD strongly correlates to traditionally used ultrasound metrics in DDH diagnostics.
- Primary PFD screening may be a viable candidate for selective screening for DDH as it predicts acetabular morphology and hip stability upon follow-up hip US.

Introduction

Multiple ultrasound (US) metrics have been proposed in screening for developmental dysplasia of the hip (DDH) with the most commonly used being the alpha angle proposed by Graf in 1983 [1], which describes the morphological conditions of the acetabulum, by measuring the slope of the acetabular roof. In the following years Harcke and Terjesen proposed the femoral head coverage (FHC) which is defined as the percentage of the cartilaginous femoral head covered by the bony acetabular roof while the hip is at rest and while applying lateralizing stress [2 3].

Both clinical stability testing and FHC during provocation evaluates the degree of laxity of the paediatric hip joint. To avoid any confusion in terms, we have chosen to use the term "displaceability" when referring to hips able to be provoked laterally during hip US i.e. a FHC < 50% in situ or during provocation while "instability" refers to clinical instability.

In 2013, the Pubo-Femoral Distance (PFD) US method was first published [4]. PFD measures the minimum distance between the medial epiphysis of the femoral head and the ossified pubic bone while applying lateralizing stress to the hip joint. It is a stress similar to the FHC but rather than being measured in relative units (percentages) it is measured in millimetres and thus does not account for individual differences in the size of the examined anatomy. The PFD has been proven to be a reliable measure and an accessible alternative to traditionally used US metrics[5 6]. The original authors have already implemented universal screening of female newborns in their region of France using the PFD method, which has reportedly reduced the rate of late diagnoses of hip dysplasia to zero in a catchment area of one million inhabitants[4]. However, the diagnosis of DDH was not made using the gold standard Graf method, rather it relied on an assessment of clinical stability and acetabular morphology in US using the PFD measurement and FHC[4].

No studies have examined how the PFD relates to traditionally used US metrics in DDH diagnostics, including the gold standard Graf method. The aim of the current study is therefore to evaluate the correlation of PFD to alfa angles and hip displaceability, as measured by the FHC in rest and during lateralizing stress, in newborns undergoing US screening for DDH.

Methods

Design and setting: This was a retrospective observational study of newborns referred for DDH US screening at INSTITUTION, COUNTRY, during a one-year period from October 2021 to October 2022. Annually, 5000 newborns are born at INSTITUTION a tertiary hospital including the only maternity ward in the municipality of CITY. Reporting follows the STROBE guidelines for reporting on observational studies [7].

Participants

The newborns in the present retrospective study participated in the Danish Hip Screening Project (DHP). In the DHP, primary early PFD screening was added to the traditional selective referral criteria for follow-up Graf hip US. A newborn was included in the DHP, and the present retrospective study, once written parental consent for participation and data collection had been obtained.

The clinical examination and risk factor identification were performed by a midwife at the post-partum clinic at INSTITUTION. The primary PFD US examination was performed by a secondary midwife trained in the PFD method on the same weekday or, in the case of the newborn being screened in the post-partum clinic in the weekend, in the following week. All examinations including clinical-, risk factor- and primary PFD screening were performed within the first 10 days after birth.

We included newborns referred in this hybrid selective screening program for DDH at INSTITUTION where primary clinical examination, risk factor identification and primary PFD US examination had been performed. Referral criteria were: a positive clinical examination, presence of a risk factor (family history of DDH, breech presentation, oligohydramnios, twins, clubfeet or musculoskeletal syndromes) a primary PFD $\geq$ 5.1 mm or a PFD difference $\geq$ 1.5mm between hips.

The exclusion criteria were: newborns with musculoskeletal syndromes, age at follow-up US examination by radiologist above three months, follow-up US examination missing PFD measurements.

The referred newborns received a follow-up hip US examination ideally before six weeks of age or, in the case clinical instability or a primary PFD above 8.0, before two weeks of age. Follow-up hip US was performed by one of three musculoskeletal radiologists experienced in paediatric US using a combination of the Graf, Harcke and PFD methods [1 2 4 8] (figure 1). Alpha angles and FHC were measured in the frontal standard plane with the child fixed in a cradle in the lateral decubitus examination position, the hip flexed to 90 degrees and the knees gently adducted. Secondly, the FHC during provocation and PFD was measured in the same lateral examination position while the hip was stressed laterally in a Barlow equivalent manoeuvre. FHC during provocation was routinely reported in all scans by one of three radiologists, while the remaining two only reported it, if they observed lateralization on US i.e. FHC decreased below 50% after applying lateralizing stress to the hip.

All measurements were performed using a high frequency (10 MHz) linear transducer (Model: Canon Aplio i800; Canon Medical Systems, Tokyo, Japan). The parents were present during all examinations of the newborns.

Statistical methods

Variables: the performed analyses use alpha angles, FHC and FHC during provocation as dependent variables and PFD as obtained at the follow-up hip US examination as dependent variable. The primary PFD measurement was only used in the referral of patients, not for subsequent correlation analysis.

We examined the correlation and impact of increasing PFD values on alpha angles, FHC, and FHC during provocation using linear regression, scatter plots, and box plots. Regression results are presented as intersections and β-coefficients with accompanying p-values. Scatter plots are presented with fitted lines and 95% confidence intervals and linear regression coefficients. To further illustrate the correlation, mean PFD values were calculated stratified by Graf classification and hip displaceability status and compared using Student's t-test as well as box plots with median values and 25% and 75% centiles with whiskers representing upper- and lower adjacent values. A sensitivity analysis was performed using a mixed effect model to account for any bias introduced by the bilaterality of observations. As no bias was detected in the sensitivity analysis, independence between bilateral observations was assumed. PFD measurements were used as a referral criterion, when performed by a midwife in primary screening, and as independent variable in the regression analysis, when performed by the radiologist at the follow-up hip US. To investigate any selection bias introduced to the correlation analysis of PFD to alpha angles and displaceability status, by selecting patients using primary PFD screening, a secondary sensitivity analysis was performed by linear regression stratified by referred/not referred by primary PFD screening. No significant difference in regression coefficients between these two groups was detected. Normality of data was inspected using QQ-plots for continuous data and a significance level of 5% was used.

All statistical analyses were performed using Stata version 17.0 (StataCorp, College Station, TX, USA).

Ethics

Ethical approval for the present study and written parent information was obtained from the COUNTRY National Committee on Health Research Ethics (Registration number N-20200051).

The project and data management plan were approved by the regional Department of Research Data and Statistics at INSTITUTION (Project ID 2021-043).

Results

In the present study 4,794 newborns were born during the study period. Of these, consent for data collection was obtained from the parents of 2,735 newborns. 815 newborns were referred for follow-up hip US, 53 didn't show, eight were referred to another institution, six hips had no PFD US measurements and two newborns were older than three months at US examination, which left 752 newborns for inclusion (1500 hips) (Figure 2). Gender distribution was 372 males and 380 females, mean age at examination was 36.6 days (range 4-87 days). Distribution of patients according to highest Graf classification was type I: 696 (92.5%), type IIa: 48 (6.4%), type IIc: 7(0.9%) and type III: 1(0.1%). Distribution of hips according to Graf classification were type I: 1.416, type IIa: 74, type Iic: 9 and Type III: 1. 78 hips were classified as displaceable and 1422 were non-displaceable (table 1).

Inspection of scatter plots and linear regression revealed a negative linear correlation of PFD to alpha angles ($p<0.001$), FHC($p<0.001$) and FHC during provocation ($p<0.001$) with a 1mm increase in PFD corresponding to a -2.1 degree (95% CI -2.3;-1.9) change in alpha angle, a -3.4% (95% CI -3.7;-3.0) change in FHC and a -6.0% (-6.6;-5.5) change in FHC during provocation (Table 2, figure 3, figure 4). Further, PFD was significantly higher with increasing Graf types and in displaceable hips ($p<0.001$) (Figure 5, Table 1).

Discussion

Key results

PFD was significantly correlated to both acetabular morphology and hip displaceability. An increase in PFD was seen with both shallowing of the acetabulum and an increase in hip displaceability.

Interpretation

There is no universal consensus on what constitutes true DDH. Graf proposed a treatment plan according to his classification system which relies on hip morphology [9]. Surgeons, when deciding which hips to treat, rely on a combination of radiological findings and hip stability assessment, with the latter being the guiding factor for a majority of surgeons, as they are more likely to opt for nonoperative management of children showing no signs of hip instability [10]. Hip instability is clinically assessed using the Barlow and Ortolani manoeuvres, and the Galeazzi test, but the value of these examinations is questionable. The Barlow and Ortolani manoeuvres have a combined sensitivity of 60% [11]. While they are more sensitive in the hands of experienced orthopaedic surgeons [12], in a study from 2020, Harper and colleagues demonstrated that 14% of dislocated hips, as detected on US, were incorrectly classified as being reduced upon clinical examination by experienced orthopaedic surgeons [13]. Further, the positive predictive value of clinical hip examinations in detecting hip dysplasia defined as ≥Graf IIc type hips, is 33% among experienced orthopaedic surgeons [14] and as low as 4% among primary screeners [15].

In evaluating the correlation of PFD to hip stability, we therefore chose to define it as displaceability using the FHC which has a high degree of agreement when classifying dysplastic hips (Kappa>0.7)[16]. Terjesen originally described a cut of value for FHC of 44% for females and 47% for males for hip dysplasia with some age variation. Others describe a cut-off of 50% for both genders [17], which is also used routinely at INSTITUTION and consequently in this study.

In terms of reliability and accessibility, the PFD method outperforms both the Graf- and FHC methods [16 4 5] , but, as demonstrated in this study, is strongly correlated to both.

The PFD method may therefore be a viable candidate for predicting alpha angles and displaceability status in primary DDH screening.

Limitations

Both the PFD and FHC methods rely on a Barlow equivalent hip provocation manoeuvre. As the application of force may not be equal between the examiners when

performing the examinations, the obtained measurements may to a minor extent be affected in precision. This uncertainty can affect the precision of the correlation coefficients of our regression analyses. However, the impact may not have any clinical influence when classifying the hips as displaceable using the 50% FHC threshold as a significant increased PFD when compared to stable hips was found.

The present study only evaluated the correlation of PFD to Graf's alpha angles and hip displaceability status. An assessment of the effectiveness of primary PFD screening on DDH detection cannot be made on the present results.

Generalizability

Participants for the current study were selected through a unique selective screening programme for DDH using primary PFD as a referral criterion for follow-up Graf hip US. As such, all newborns included in this study were selected based on the presence of a risk factor for DDH, positive clinical examination or primary PFD screening. However, as the prevalence of Graf types are comparable to those reported in universal screening programmes [18], we believe the present results to be representative of a general population.

Conclusion

PFD is strongly correlated to both alpha angles and hip displaceability, as measured by FHC and FHC during provocation, in US of hips at six weeks of age. The PFD increases as the hips become more dysplastic and/or displaceable.

References

1. **Graf R**. New Possibilities for the Diagnosis of Congenital Hip Joint Dislocation by Ultrasonography. Journal of Pediatric Orthopedics, 1983.

2. **Harcke HT, Clarke NM, Lee MS, Borns PF, MacEwen GD**. Examination of the infant hip with real-time ultrasonography. *J Ultrasound Med* [Internet] 1984;3(3):131–7.

3. **Terjesen T, Bredland T, Berg V**. Ultrasound for Assessment in the. *J Bone Jt Surg [Br]* 1989;71(5):767–773.

4. **Tréguier C, Chapuis M, Branger B, Bruneau B, Grellier A, Chouklati K, et al.** Pubo-femoral distance: An easy sonographic screening test to avoid late diagnosis of developmental dysplasia of the hip. *Eur Radiol* 2013;23(3):836–844.

5. **Husum H-C, Bach Hellfritzsch M, Maimburg RD, Henriksen M, Lapitskaya N, Møller-Madsen B, et al.** Pubo-Femoral Distances Measured Reliably by Midwives in Hip Dysplasia Ultrasound. *Children* [Internet] 2022;9(9):1345.

6. **Teixeira SR, Dalto VF, Maranho DA, Zoghbi-Neto OS, Volpon JB, Nogueira-Barbosa MH**. Comparison between Graf method and pubo-femoral distance in neutral and flexion positions to diagnose developmental dysplasia of the hip. *Eur J Radiol* [Internet] Elsevier Ireland Ltd, 2015;84(2):301–306.

7. **Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP**. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61(4):344–349.

8. **Couture A, Baud C, Prodhomme O, Saquintaah M, Veyrac C**. Ultrasound of the neonatal hip: initial evaluation and follow-up. *J Radiol* [Internet] 2011;92(2):142–165.

9. **Graf R**. *Hip Sonography. Diagnosis and Management of Infant Hip Dysplasia*. 2nd ed. Springer. Springer, 2006.

10. **Ömeroğlu H, Ağuş H, Biçimoğlu A, Tümer Y**. Evaluation of experienced surgeons' decisions regarding the need for secondary surgery in developmental dysplasia of the hip. *J Pediatr Orthop* 2012;32(1):58–63.

11. **Jones D**. An assessment of the value of examination of the hip in the newborn. *J Bone Joint Surg Br* [Internet] 1977;59(3):318–22.

12. **El-Shazly M, Trainor B, Kernohan WG, Turner I, Haugh PE, Johnston AF, et al.** Reliability of the Barlow and Ortolani Tests for Neonatal Hip Instability. *J Med Screen* [Internet] 1994 [cited 20 Jan 2020];1(3):165–168.

13. **Harper P, Joseph BM, Clarke NMP, Herrera-Soto J, Sankar WN, Schaeffer EK, et al.** Even Experts Can Be Fooled: Reliability of Clinical Examination for Diagnosing Hip Dislocations in Newborns. *J Pediatr Orthop* Lippincott Williams and Wilkins, 2020;40(8):408–412.

14. **Husum H-C, Ghaffari A, Rytoft LA, Svendsson J, Harving S, Kold S, et al.** Positive predictive values in clinical screening for developmental dysplasia of the hip. *Acta Paediatr* [Internet] Norway, 2021;110(8):2430–2434.

15. **Choudry QA, Paton RW**. Neonatal screening and selective sonographic imaging in the diagnosis of developmental dysplasia of the hip. *Bone Jt J* 2018;100B(6):806–810.

16. **Quader N, Schaeffer EK, Hodgson AJ, Abugharbieh R, Mulpuri K**. A Systematic Review and Meta-analysis on the Reproducibility of Ultrasound-based Metrics for Assessing Developmental Dysplasia of the Hip. *J Pediatr Orthop* [Internet] United States, 2018;38(6):e305–e311.

17. **Ömeroğlu H, Õmeroğlu H**. Use of ultrasonography in developmental dysplasia of the hip. *J Child Orthop* [Internet] 2014;8(2):105–113.

18. **Biedermann R, Riccabona J, Giesinger JM, Brunner A, Liebensteiner M, Wansch J, et al.** Results of universal ultrasound screening for developmental dysplasia of the hip. *Bone Joint J* [Internet] 2018 [cited 17 Jan 2020];100-B(10):1399–1404.

Tables and figures

Graf classification (n hips)	PFD (mean 95% CI)	P-value
Type I (n=1,434)	3.8 (3.8;3.9)	
Type IIa (n=74)	5.4 (5.1;5.7)	<0.001*
Type IIc (n=9)	7.4 (6.4;8.5)	<0.001*
Type III+ (n=1)	11.2 (N/A)	N/A
Hip displaceability		
Non-displaceable = FHC>50% (n=1,440)	3.8 (3.8;3.9)	
Displaceable= FHC <50% (n= 78)	5.9 (5.6;6.2)	<0.001*

Table 1: distribution of hips according to Graf types and hip displaceability status with accompanying mean PFD values. PFD = Pubo-Femoral distance

Variable	Intersection (crude)	β-coefficient (95% CI) (crude)	P-value
Alpha	74.5°	-2.1 (-2.3;-1.9)	<0.001*
FHC	76.0%	-3.4 (-3.7;-3.0)	<0.001*
FHC with provocation	84.1%	-6.0 (-6.6;-5.5)	<0.001*

Table 2: Results of linear regression of alpha angles, FHC and FHC during psrovocation with Pubo-Femoral distance as independent variable. FHC = femoral head coverage.

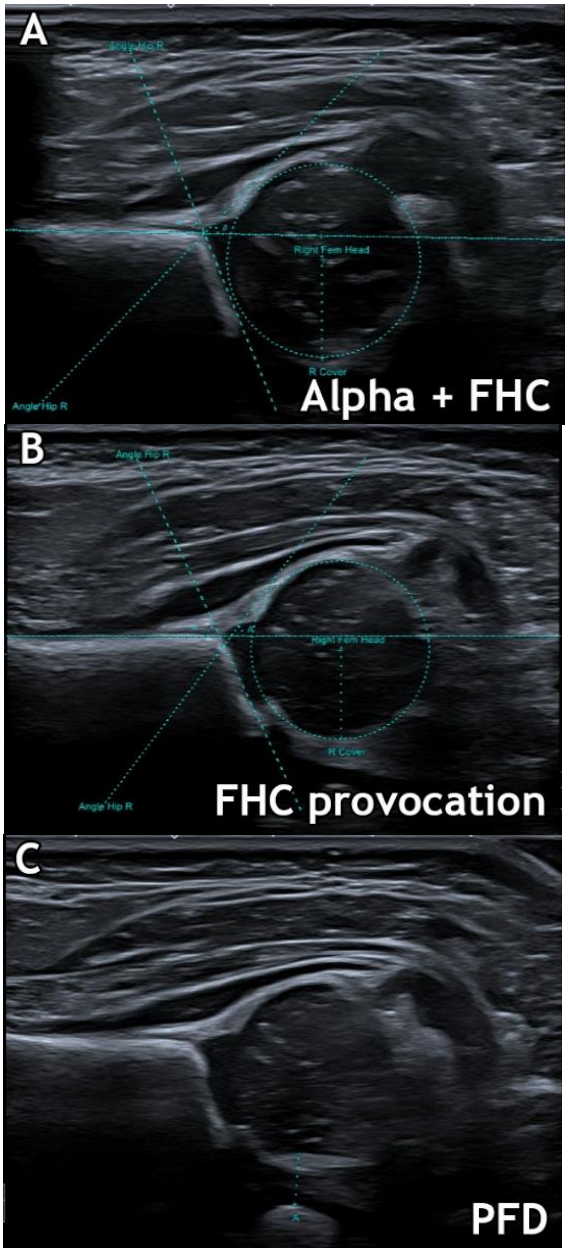


Figure 1: Three ultrasound images of one paediatric hip examination. A) Graf standard plane with added annotated alpha angles and FHC values. B) ultrasound image captured during hip provocation with annotated FHC values. C) Ultrasound image captured during hip provocation with annotated PFD values. PFD = Pubo-femoral distance, FHC = Femoral Head Coverage.

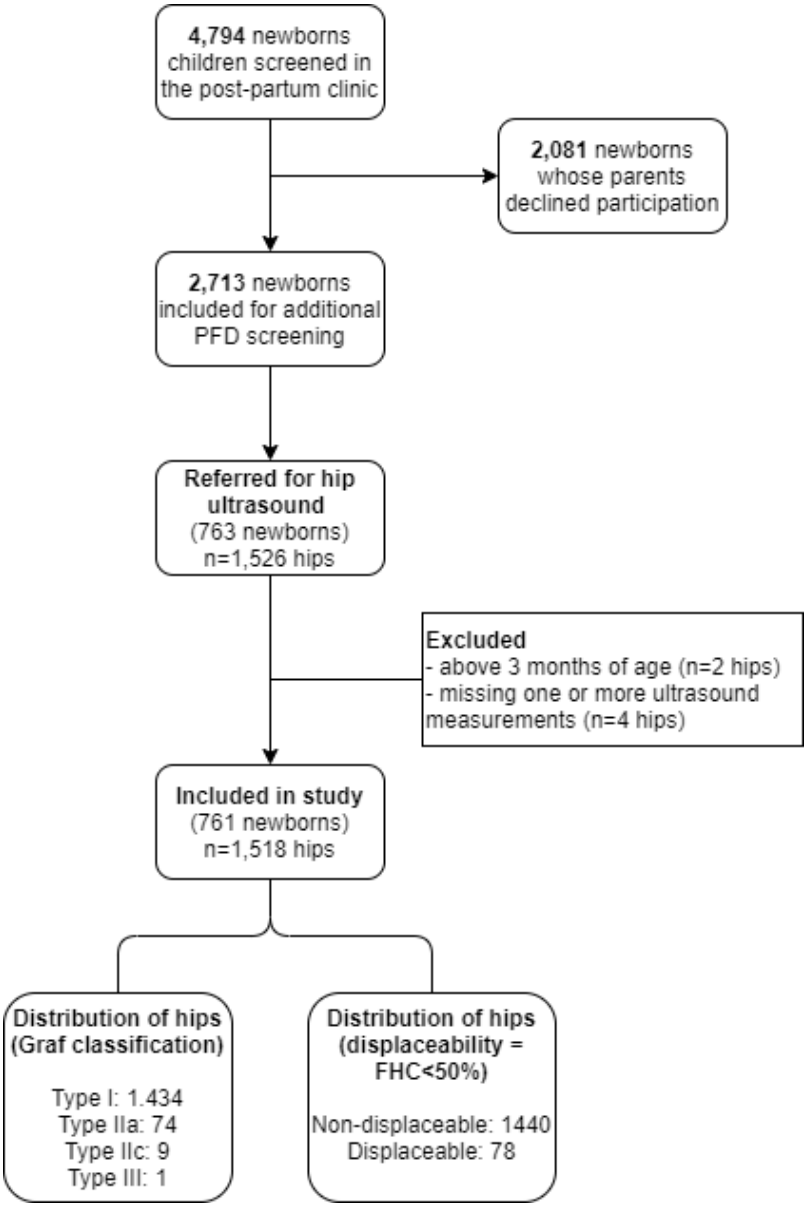


Figure 2: Consort diagram of inclusion process and distribution of hips according to the Graf classification and displaceability criteria. FHC = Femoral Head Coverage

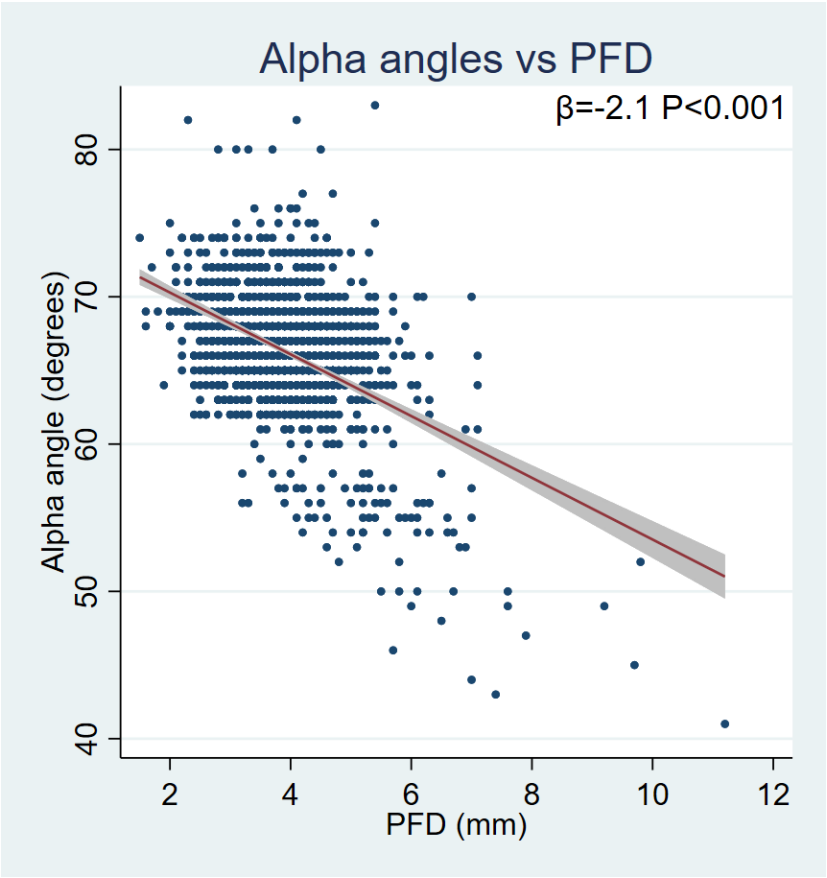


Figure 3: Scatter plot of PFD and Alpha angles with fitted regression line, 95% confidence intervals and regression coefficients. PFD = Pubo-Femoral distance.

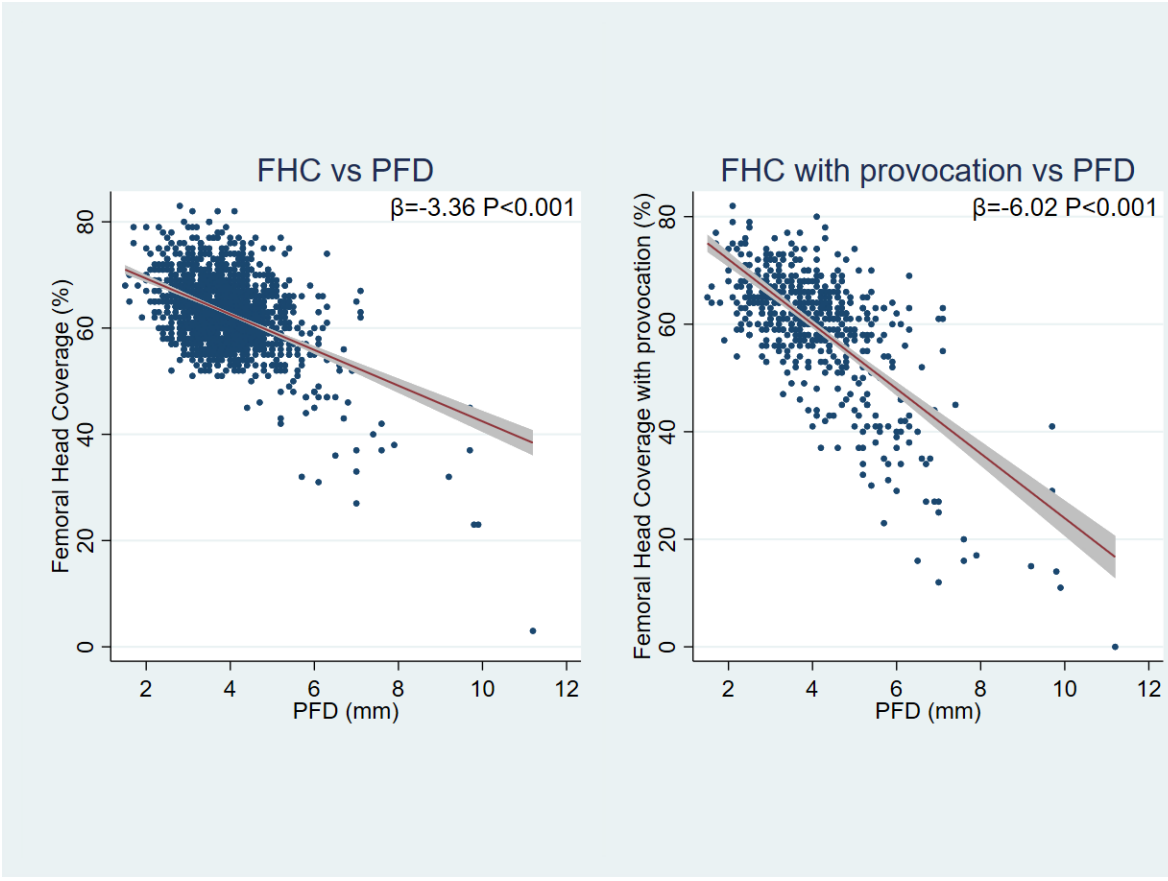


Figure 4: Scatter plot of PFD, FHC and FHC with provocation with fitted regression lines, 95% confidence intervals and regression coefficients. PFD = Pubo-Femoral distance, FHC = Femoral Head Coverage.

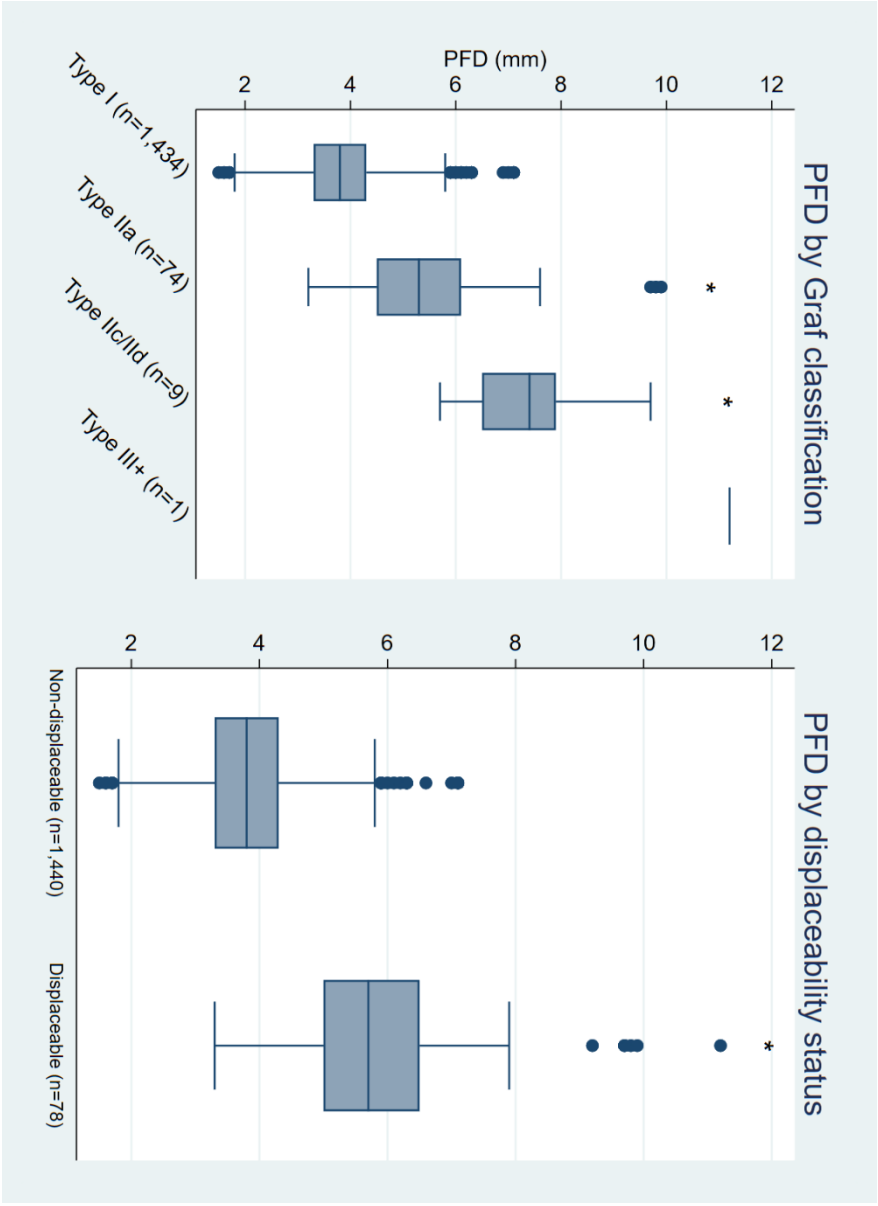


Figure 5: Box plots of PFD values stratified by Graf classification and hip displaceability status. Boxes represent 25%, median and 75% percentiles with whiskers representing upper- and lower adjacent values. PFD = Pubo-Femoral Distance. Displaceable = FHC or FHC during provocation <50% *=significant result. FHC = Femoral Head Coverage.