Three-dimensional ultrasound in the monitoring of hip dysplasia in cerebral palsy

Kay, Becky

Awarding institution:
King's College London

Three-dimensional ultrasound in the monitoring of hip dysplasia in cerebral palsy

Rebecca Hazell Kay

2021

School of Biomedical Engineering & Imaging Sciences

King's College London

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Acknowledgements

The work presented in this thesis would not have been possible without the contributions, advice, guidance, and encouragement of the following people.

Firstly, I would like to thank my supervisors, Dr Adam Shortland and Professor Stephen Keevil. Adam, you have provided direction, insight, vision, enthusiasm, and encouragement from the very beginning right through to the completion of this thesis. Steve, thank you for your knowledge, guidance, and time throughout this PhD. Thank you also to Professor Lionel Tarassenko, Dr Anne McNee, Dr Robert Eckersley and Dr Jorge Cardoso for your time, challenges, advice, and expertise throughout.

I would like to thank Action Medical Research and all those who kindly donated to this project, I hope that the work in this project contributes to the continued efforts to advance and improve the clinical outcomes of children with cerebral palsy.

Mr Gough, Mr Kokkinakis and Mr Reed thank you for your invaluable insights and commitment to clinical study, without your help and dedication the study would not have been possible.

Thank you to Jonathan Noble, Liam Johnston and Mike Jeffryes for your inputs to both the clinical and *in vitro* studies, your suggestions and contributions are very valued. Thank you also to the wider team at the One Small Step Gait Laboratory for providing such a stimulating and enjoyable workplace for the duration of this project.

To the many research nurses and staff at the Clinical Research Facility, thank you for enabling and facilitating the clinical study.

Thank you to Mark Stanton, Bill Stevens and Steve Turner for your expert assistance in the manufacture of the mechanical rig and to Nick Byrne for all your time and help in the printing of the 3D hip phantoms.

I would also like to thank my husband Chris and my family for your patience, continued encouragement, interest, and distraction when it's most needed.

Finally, a big thank you to the amazing volunteers and their families who gave up their time to participate, without your support this project would not have been possible.

Contents

Figures

Tables

Equations

Abbreviations

CP	Cerebral Palsy
CPUP	The national CP surveillance programme in Sweden
GMFCS	Gross Motor Function Classification System
QOL	Quality of Life
CHQ	Childhood Health Questionnaire
HRQL	Health Related Quality of Life
RMP	Reimer's Migration Percentage
AI	Acetabular Index
HSA	Head-shaft angle
AP	Anterior-Posterior
HEA	Hilgenreiner's epiphyseal angle
NSA	Neck-shaft angle
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
3DUS	Three-dimensional Ultrasound
ICC	Intraclass Correlation Coefficient
CAPS	Chailey Adjustable Postural Support
BTX-A	Botulinum Toxin A
AACPDM	American Academy of Cerebral Palsy and Childhood Disability
VDRO	Varus de-rotation osteotomies
MDD	Minimal detectable difference
RF	Radio frequency
DDH	Developmental dysplasia of the hip
ROM	Range of motion
LHD	Lateral Head Distance
TP	True positive
FP	False positive
TN	True negative
FN	False negative
PPV	Positive predictive value
NPV	Negative predictive value
QALY	Quality-Adjusted Life Years
WHO	World Health Organisation

SEM	Standard error of measurement
LHC	Lateral Head Coverage
FHPPR	Femoral Head Posterior Position Ratio
FH	Femoral Head
HRA	Health Research Authority
CRF	Clinical Research Facility
GMFM	Gross Motor Function Measure

Thesis objectives

The primary purpose of this work was to investigate the validity of using 3D ultrasound to monitor hip dysplasia in children with cerebral palsy. The thesis starts with an in-depth review and critique of current clinical methods for identification of hip dysplasia, as well as setting out background knowledge and context required to understand and critique the work presented. I then investigated, using simulation, the impact of measurement uncertainty within a typical hip surveillance programme on the clinical management of a child. Collectively these chapters highlight the motivation for the investigation and development of a novel index (lateral head coverage (LHC)) derived from 3D ultrasound data to assess hip dysplasia in this population. An *in vitro* study was conducted to assess the performance of this index as well as another novel index, designed to quantify the posterior displacement of the femoral head relative to the acetabulum, (femoral head posterior position ratio (FHPPR)). Finally, a clinical study was conducted which included 25 children with cerebral palsy. Initially, the agreement between the clinical standard measurements taken from X-ray acquired in the routine care of these patients, and LHC from 3D ultrasound was investigated. The measurement of FHPPR was also taken on the images acquired as part of the study, however there was no conventional 3D imaging of the hips in these children to compare these results too. Finally, I draw together the findings from the simulation, *in vivo* and *in vitro* studies and suggest the impact that these findings might have on our understanding, monitoring and care of hip dysplasia in children with cerebral palsy. This work is a contribution to progress and requires both the clinical and scientific communities to challenge, replicate and extend the studies described; two of the four

studies have been published and are therefore visible to the community for discussion.

Figure 1 summaries this structure of this thesis for reference.

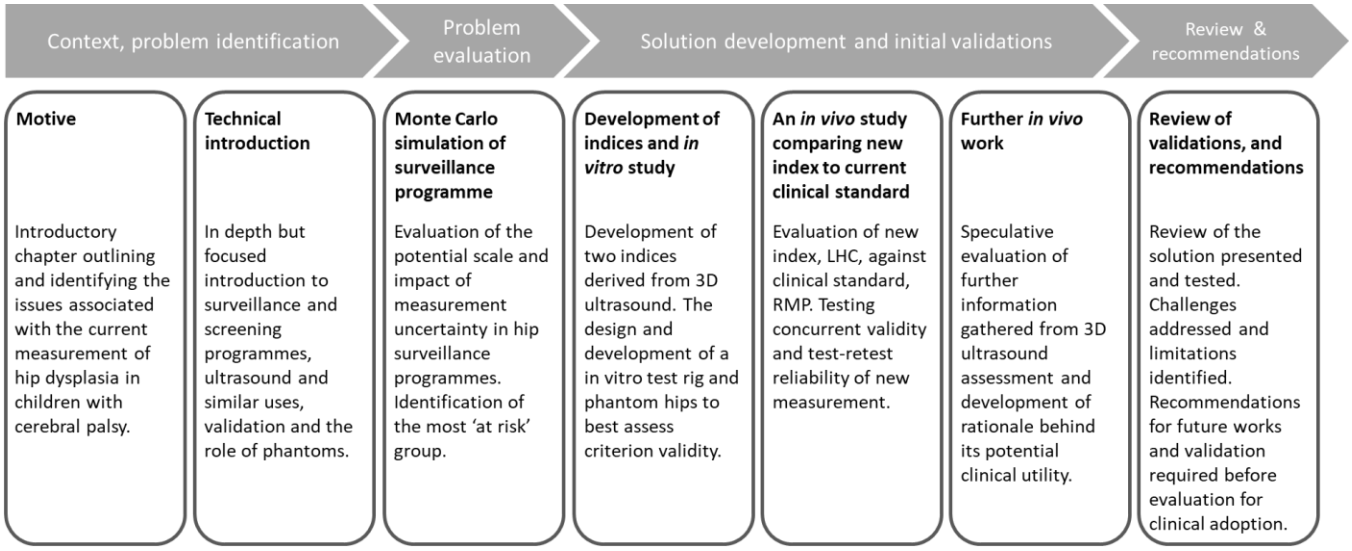


Figure 1: Summary of thesis structure and content

1. Hip dysplasia in cerebral palsy - pathophysiology, aetiology, measurement, and treatment.

1.1. Overview

This thesis concerns the evaluation of a new method of measuring hip dysplasia in children with cerebral palsy (CP). In this chapter, I review the literature pertaining to the hip in CP with a particular focus on the morphological features that have been used to quantify its maldevelopment in this group.

Cerebral palsy is defined as a “disorder of movement and posture due to a defect or lesion of the immature brain.”[1] These injuries affect motor development and can give rise to musculoskeletal deformity. One of the most common musculoskeletal problems is hip dysplasia. Whilst there is not universal agreement on a definition for hip dysplasia, Musielak et al describes hip dysplasia as the “abnormal growth of the hip” and “refers not only to the osseous structures, but also other tissues (including soft tissues) forming the structure of the hip”[2]. To aid with describing particular presentations, more specific definitions have been developed. Hip subluxation describes an “incomplete dislocation with incomplete contact between articular surfaces of the acetabulum and the femoral head”. Dislocation describes a “loss of contact between the articular surfaces of the acetabulum and the femoral head”[2]. If left un-diagnosed and un-treated there is a risk that hip subluxation can develop to hip dislocation. Hip subluxation is also commonly described as hip migration or hip displacement, particularly when describing hip dysplasia within the cerebral palsy population. Hip migration comprises of the displacement of the femoral head relative

to the acetabulum. The deformation of the femoral head or acetabulum is also increasingly common as hip migration increases, and is encapsulated by the broader definition of hip dysplasia[2].

There is a strong correlation between the magnitude of hip dysplasia and the level of disability in cerebral palsy with the more-affected children having a greater risk of clinically-significant hip migration. However, the pathophysiological mechanisms leading to hip displacement have not been fully elaborated.

Treatments range from postural management to osteotomies of pelvis and femur. Surgical treatment, when performed in a timely manner, can prevent hip dislocation and reduce pain. A significant challenge in the management of hip dysplasia is identifying dysplasia, monitoring the hips of children with cerebral palsy and predicting the trajectory of hip displacement. To address this hip surveillance programmes have been developed internationally. As standardisation of clinical management increases, both regionally, nationally and internationally, so too does the need to understand the limitations of the underpinning clinical measurements.

1.2. Cerebral Palsy

Cerebral palsy (CP) is an umbrella term describing conditions arising from a non-progressive neurological lesion acquired in foetal development, at birth or within the first 2 years of life[1]. The incidence of CP varies from 1.5 to over 4 per 1000 individuals, depending on global geographical location and whether the prevalence is reported as a percentage of live births of a specified age range[3–5]. In the UK the incidence is approximately 2.2 for each 1000 live births making it the most prevalent childhood motor disability[6]. The neurological lesion is considered static, but children with CP often develop bone, muscle and joint deformities over time. Individuals may undergo many interventions during their childhood and early adulthood to correct or prevent progression of their deformities, from physiotherapy and postural management systems to botulinum toxin injections and neuro-orthopaedic surgery [7,8]. Outcomes from these interventions are often moderate, in part, due to the challenges presented by the variation in the clinical presentation of these children. The Gross Motor Function Classification System (GMFCS) was created in 1997 by Palisano *et al*[9] to aid in classifying individuals with CP by severity (Figure 2[10]). GMFCS level I describes individuals with the highest functional ability, and GMFCS level V describes individuals with the most significant mobility limitations.

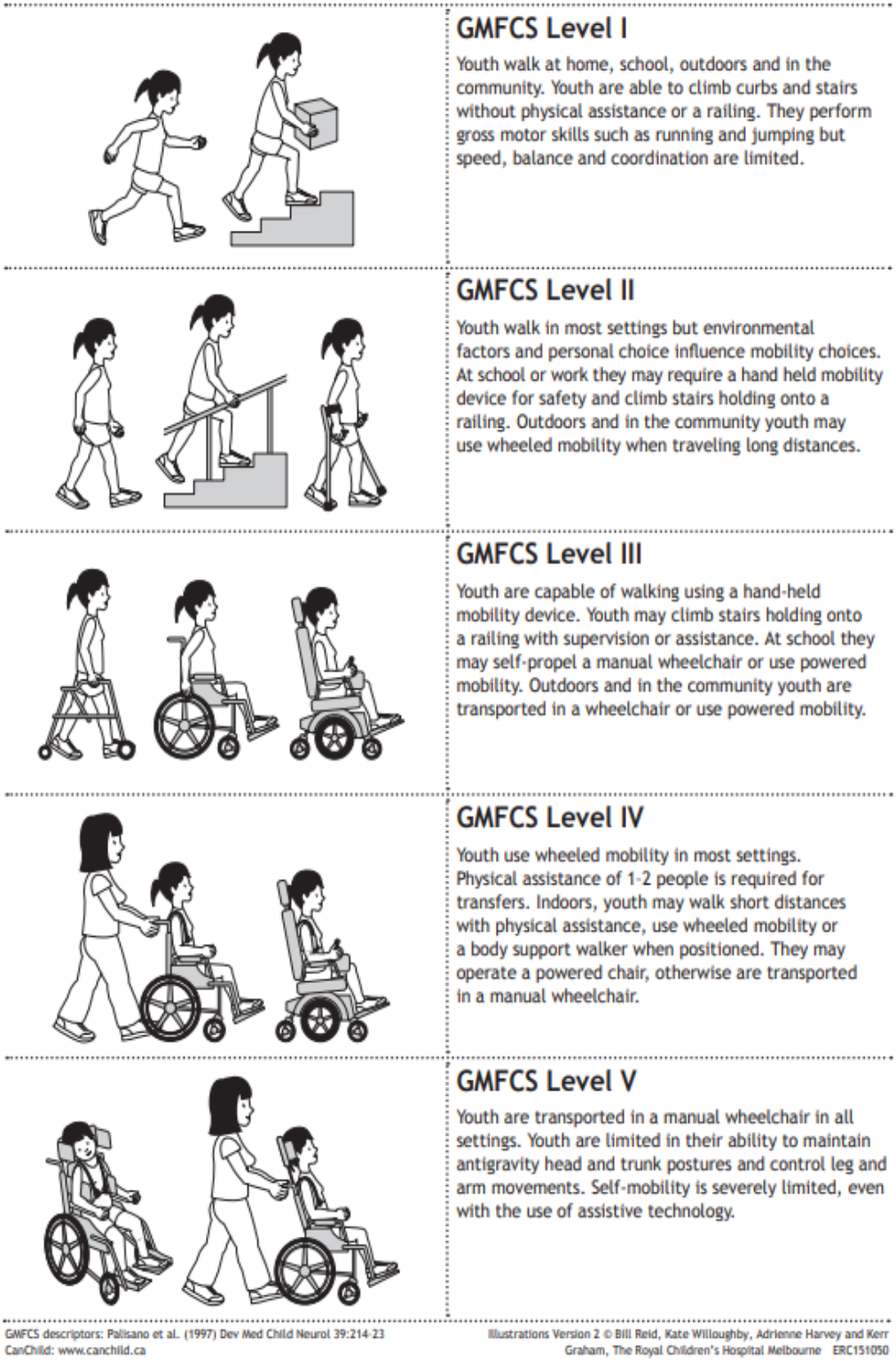


Figure 2: Illustration of GMFCS level classifications for children ages 6-11 years of age[9]

1.3. Measuring Impact and Pain

Pain is inherently hard to quantify in an objective way, even more so in populations where self-reporting is not always possible. The prevalence of pain in children with CP is variably reported to be 30%-70%[11–13] making it one of the most frequently reported secondary conditions in CP. Increasing age and female gender are risk factors for increasing pain intensity[12]. The presence of pain has an impact on quality of life, participation in daily activities and family stresses[12]. Eriksson *et al*[12] conducted a cross-sectional evaluation of pain. They reviewed their national register (CPUP) which included 3545 children aged 4-18years at the time of study. Pain was reported in 42.5% of individuals. The lowest prevalence was found in the youngest cohort, aged 4-5 years, with just under one third reporting pain. This increased to 57.3% for those aged 18 years old. There was no significant difference in the prevalence of pain across GMFCS levels I to IV, however the severity and site of pain differed between these groups. Pain intensity increased with GMFCS level, and hip/thigh pain was the most prevalent cause of pain in non-ambulant individuals. The most severe pain was also reported at hip/thigh and abdominal sites. Adolescents at GMFCS V reported the most pain.

Bagg *et al*[14] reported that hip positioning had a significant impact on severity of pain, with dislocated hips being significantly more painful than displaced hips. Pain intensity was given a rating from 1-3, with 1 being no pain and 3 being severe pain. The mean pain score within the dislocated group was 2.2 compared with 1.7 and 1.4 for the displaced and reduced hips respectively. It is worth noting that in the methodology of this study, the average follow-up period was 19 years, range 8-30 years. All hips (N=64) were displaced at initial presentation, and at follow up approximately half

(N=31) were reduced, 9 dislocated and the rest (N=24) remained displaced. It is likely that severity of is correlated with the period of time that the hip has been displaced.

In Eriksson *et al*'s study, of those who reported pain, over 60% reported that it had significant affected their daily activities in the last 4 weeks and over a third reported that it had adversely disrupted their sleep, with those reporting pain at multiple sites and higher GMFCS levels most likely to be affected[12].

The association and impact of gross motor function and quality of life (QOL) scores are variably reported[15–18], however most investigators observed a reduction in physical summary score, general health, role–physical, parent impact–time domains with increasing GMFCS level[18]. Most scales used are generic paediatric scales such as the childhood health questionnaire (CHQ)[19]. Some domains in these scales are likely to lack relevance within this population due to the levels of disability of many individuals. For example, in the behavioural domain there are questions about lying, arguing and stealing which are rarely applicable to individuals with more severe cerebral palsy due to communication deficiencies. Even within the spectrum of cerebral palsy, the breadth of impairment in the physical, cognitive and communication domains present a challenge to defining meaningful health related quality of life (HRQL) measures for this population.

HRQL measures tend to either be discriminative (capable of distinguishing between individuals), predictive (able to estimate a future outcome) or evaluative (able to detect change over time). A measure designed to enable an understanding of QOL at a population level is unlikely to be capable of detecting change related to a specific intervention. Traditionally the success of a treatment or intervention is measured by its performance in alleviating the pathophysiological impairment, however in some

cases the perceived benefit to the child or family may be in the social domain and not an immediate or measurable improvement to the impairment of function. For this reason QOL scores or HRQL outcomes should also be considered when making intervention decisions[15]. Most HRQL scores that have been used to evaluate CP populations provide some insight into the impact of the condition on an individual and their family but their application in clinical research and treatment efficacy trials is limited by poor standardisation of scale, responsiveness of the scales to clinically significant changes and a lack of face validity in the relevant populations.

1.4. Hip anatomy

The hip joint (Figure 3) is a complex articulation that begins to develop *in utero* and continues to develop through childhood. The revolute joint allows for three rotational degrees of freedom. The femoral head is broadly spherical and congruent to the acetabulum, often referred to as the socket. The acetabular labrum is a fibrocartilaginous, horse-shoe shaped structure which lines the rim of the acetabulum. The labrum and surrounding ligaments provide stability to the hip joint. There are many

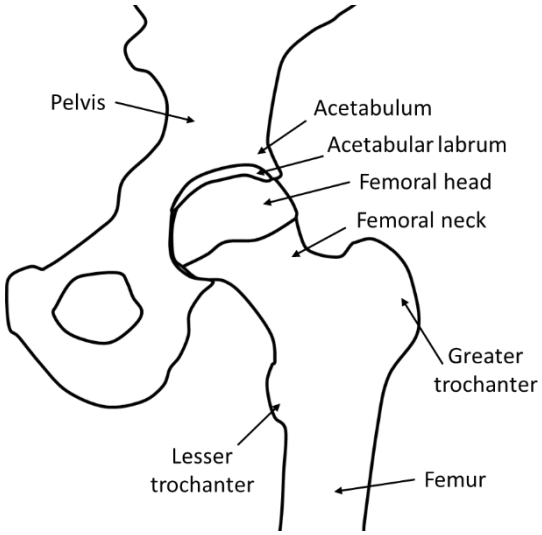


Figure 3: A schematic of the hip joint

muscles that insert around the hip joint to facilitate the wide range of motion that the joint design affords. Broadly these can be split into the hip adductors (adductor magnus, longus and brevis, gracilis, and pectineus), hip abductors (gluteus complex and tensor fasciae latae), hip flexors (iliopsoas, rectus femoris and sartorius) hip extensors (gluteus maximus, biceps femoris, semimembranosus and semitendinous) and rotators (piriformis, gemellus superior and inferior, the obturators and quadratus femoris)[20]. Muscles do not simply work in one plane, depending on the joint positioning the muscles may contribute to motion in other planes. For example, the gracilis muscle's primary function is as a hip adductor, but it can also flex and rotate the hip depending on the joint positioning, whilst also contributing to knee flexion and tibial rotation.

Typically, in CP, the hip appears normal at birth and it is only as the child grows that hip development can deviate from normal. Ossification of the hip complex is a dynamic process through early childhood that may be influenced by mechanical factors as well as endocrine factors[21]. The lateral aspect of the acetabulum is formed by cartilage that is not replaced by bone until near skeletal maturity[22]. The femoral head and greater trochanter ossify somewhat earlier, at approximately 6 months and 2-4 years respectively. In the young, the ossification of the femoral head may also be incomplete or eccentric with more bone formation in the lateral aspect of the femoral head in children with hip dysplasia (Figure 4).

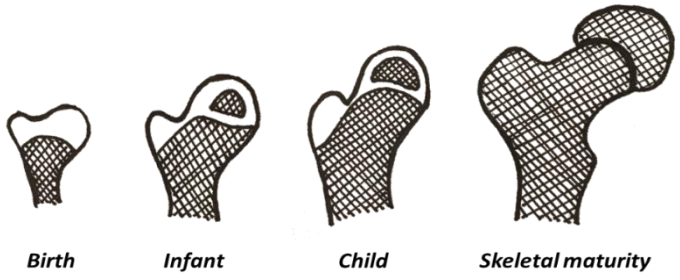


Figure 4: A schematic of the timeline and distribution of cartilage (white) and bone (hashed) in the femoral portion of the hip, adapted from Osborne et al[21]

1.5. Hip dysplasia mechanism

Hip dysplasia, in CP, is characterised by predominantly lateral displacement of the femoral head, [21]. Hip displacement can be progressive and, if left unmonitored and untreated, can lead to hip dislocation[23,24]. However, the direction of displacement is variable. Brunner *et al*[25] used computed tomography to investigate 24 hips, from 20 patients, which had fully dislocated. In the reconstructed images they discovered that in all cases a clear channel, or groove, with a unilateral direction was present in the acetabulum, which they concluded the femoral head had slid out from. The mean direction of the channel was nearly purely lateral (3° posterior to the lateral plane) but the range of directions of the channels was from 33° anterior to the lateral plane to 70° posterior to the lateral plane, indicating a significant amount of variation in the hip dysplasia mechanics. However hip dysplasia also describes deformity of the hip complex, deformity of the femoral head is common, particularly in displaced hips, and acetabular dysplasia describes the deformation of the acetabulum. Commonly, presenting as a shallower socket with more rounded edges.

It is thought that muscle imbalance around the hip and abnormal loading forces influence the abnormal development of the hip in individuals with hip dysplasia[26,27]. It is interesting that within the CP population, the children with reduced ambulatory function are more susceptible to mal-development of the hip joint than those with the ability to independently ambulate [28–31]. The favoured theory is that increased tone (increased tension in the muscles preventing relaxation) in the hip adductors is causal. Certainly within the subtypes of CP, individuals with spastic quadriplegia are the most at risk of hip dysplasia[32,33]. However there are children with cerebral palsy who are hypotonic who may go on to develop hip dysplasia[33]. Since these individuals do not

have increased tone in their hip adductor muscles, this suggests that other factors influence hip development. Reduced movement and the lack of weightbearing activities undertaken by children with CP, in comparison with their typically-developing peers, is accepted to be a contributory factor.

One alternative theory is that of persistent foetal positioning during infancy. In normal development, the femur undergoes significant remodelling in the early years. Typically, as a child develops the femoral neck angle reduces becoming more varus. The level of femoral anteversion, or proximal rotation of the femoral neck relative to the shaft of the femur, also typically reduces with increasing age. This alternative theory suggests that these changes that are typically expected do not occur or occur to a lesser extent.

1.6. Prevalence and risk factors

Hip dysplasia is a common cause of pain and disability in cerebral palsy, with a reported prevalence of up to 60%[33,34], depending on the definition and the level of involvement. After equinus deformity, a contracture of the ankle complex, it is the second most common orthopaedic problem in this population[35]. Dislocation can be prevented. This typically involves close monitoring of the hip and surgical intervention if certain levels of displacement are exceeded. The risk factors for hip dysplasia include age[29–31], subtype of cerebral palsy[30], proximal femoral geometry[29] and level of motor function, with patients in Gross Motor Function Classification System (GMFCS) level V being at greatest risk[30,31,33] (Figure 5).

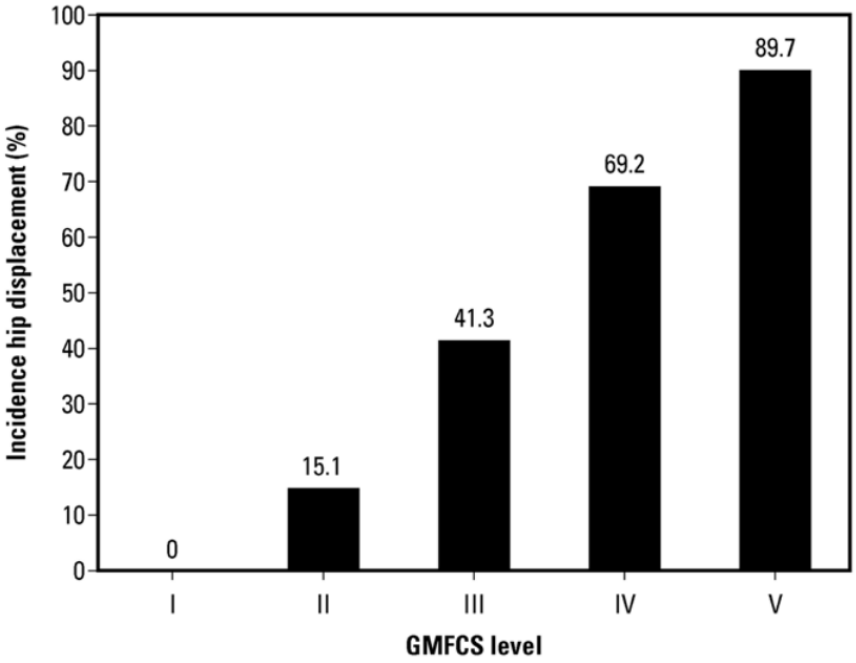


Figure 5: Incidence of hip displacement by GMFCS level, a population study in Australia[32]

Hermanson *et al*[29] created a risk score (Equation 1) to predict the likelihood of an individual with CP, GMFCS level IV or V, developing hip displacement, as measured by migration percentage >40%, within 5 years. They used an equation weighted by GMFCS level, age, and two measures from 2D radiographic imaging. The head shaft angle, which is a measure of femoral geometry, and Reimer's migration percentage (RMP) which is a measure of lateral displacement. Their sensitivity and specificity analysis showed they could differentiate between and high risk and low risk individual with an accuracy of 87%.

$$Risk\ score = -14.1 + 0.71\ (GMFCS_{IV}) + 2.48\ (GMFCS_{V}) + 0.07HSA_{max} + 0.09MP_{max} - 0.5Age$$

Equation 1: CPUP risk score developed by Hermanson et al[28] GMFCSIV is a dichotomous indicator variable which assumes the value 1 when the individual has GMFCS level IV, and 0 otherwise. MP is Reimer's migration percentage and HSA is head shaft angle (figure 6)

Terjesen published natural history data on hip dysplasia progression rates stratified by GMFCS level and sub-type of cerebral palsy[32]. He quantified hip dysplasia using RMP, and reported mean annual progression rates varying from 0.2% (s.d. 3.7%) for individuals with GMFCS level I to 9.5% (s.d. 9.4%) for individuals with GMFCS level V[32]. Park *et al*[36] reported lower progression rates of 0.3%, 1.9% and 6.2% for GMFCS levels III to V respectively. They also reported a significant difference between each of the groups supporting GMFCS level being a significant risk factor.

1.7. Identification of hip dysplasia

Hip dislocation can be prevented if hip dysplasia is detected early. However, the challenge is in correctly identifying individuals with progressive displacement as some hips do not progress, or may even improve[37]. It is often preferable to treat hip dysplasia prior to it becoming symptomatic. Current clinical practice favours radiological examination in the form of an anterior-posterior (AP) radiograph supported by passive hip range of motion examination[21,24,38,39]. The passive range of motion of the hips is measured and asymmetries or limitations in ranges are often recorded in hip assessments. They are considered indicative of asymmetric forces around the hip which in turn are thought to be at least partially causative[26,27] of hip dysplasia, however Hagglund *et al.*[30] concluded that passive range of motion around the hip was a 'poor indicator of hips at risk'. To assess the severity and quantify hip dysplasia imaging assessment is required. There are many measurements or indices used to either quantify the hip development, predict the risk of hip dysplasia or quantify the level of dysplasia[27,40,41]. Considering the reported precision of these measurements, none can really be considered a gold standard.

1.8. Indices hip dysplasia

1.8.1. Reimer's migration percentage (RMP)

Reimer's migration percentage (RMP) is the most widely used clinical measurement of hip dysplasia. RMP was initially created by Reimer in 1980[27], based on an idea of Rang's (1975). It is a simple 2D ratio of two distances measured in a single plane, describing the uncovering of the femoral head by the acetabulum (Figure 6)[27].

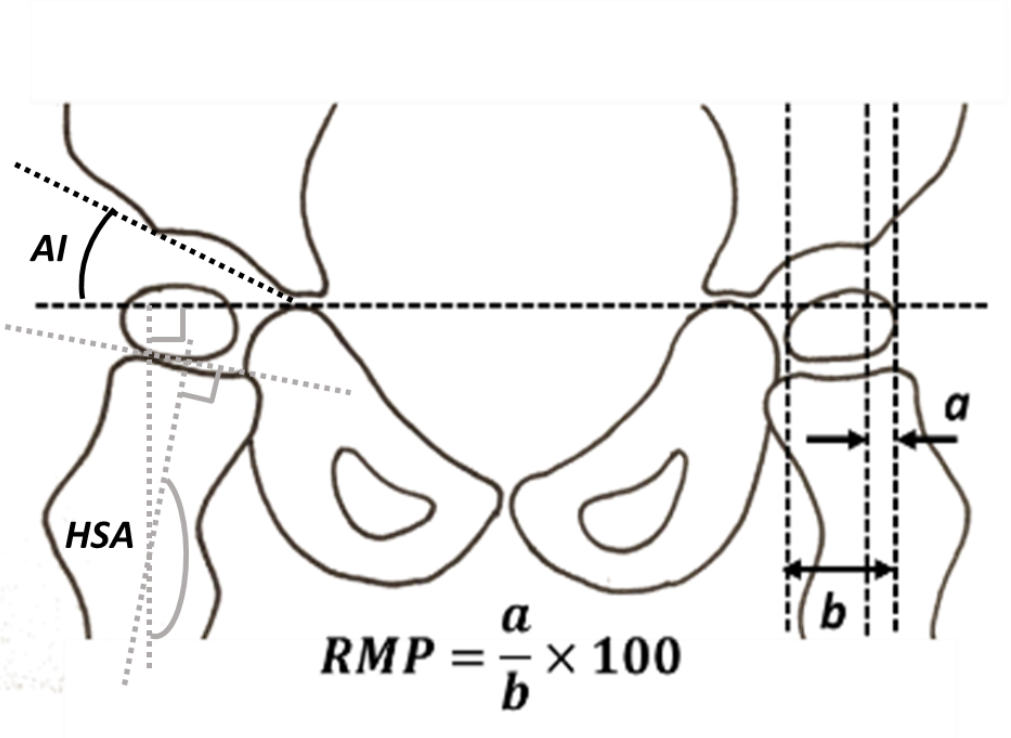


Figure 6: Measurement of Reimer's migration percentage (MP), acetabular index (AI), head shaft angle (HSA). Figure adapted from Hagglund et all[29]

Clinicians use thresholds of clinical significance to direct clinical management. However, there is not global consensus on the boundaries of the thresholds, with some questioning whether there should even be thresholds. Where thresholds are adopted, even within centres practice can vary. Hagglund et al[37] analysed their retrospective data to look at the consequence of using different RMP thresholds. In their cohort, that

spanned all GMFCS levels, approximately one third had RMP > 30%. Within this group, 88% reached the less conservative threshold (RMP > 33%) and 54% reached RMP > 40%. Of the total group with RMP > 30%, they reported that one third of the cohort decreased below the 30% threshold without operative intervention. Of the group with RMP > 40% (i.e. those that were indicated for surgery), a fifth corrected to MP < 30% without surgical intervention whilst 25% of the surgery group required a second corrective surgery. Wordie et al[42] conducted a retrospective evaluation of the Scottish registered CPIPS programme, identify a large cohort of individuals (N=453) who had multiple (at least 3 X-rays), and recorded at least one RMP reading over 35%. In this sample population they also observed 'correction', i.e. a lower RMP at a later time point without intervention, in some individuals. However, they also identified a 'point of no return'. In their population, no individual with an RMP measuring 46% or over, ever corrected to below this threshold without intervention. The team conclude that an individual whose hip migrations never reach 46% may not require interventions to relocate their hips as long as they do not have any other indications for intervention.

1.8.2. Acetabular index (AI)

Acetabular index (AI) is the most widely used clinical measurement of acetabular development. Originally documented in 1925, in German, by Hilgenreiner, and later in English by Kleinberg and Lieberman[40] the AI was developed as a measure for screening new-borns for developmental dysplasia of the hip (DDH). It provides a measure of the inclination of the acetabular roof, with a lower number indicating a more curved, or deeper, acetabulum. The index is defined as "the angle formed between the roof or iliac portion of the acetabulum and a horizontal line passing

through the triradiate cartilages" (Figure 6). Kleinberg and Lieberman's paper states that in normal new-borns the AI was 27.5° which decreases to 20° at 2 years of age. In 1989 Cooke et al[23] found AI to have the best predictive value of any single radiographic measurement for predicting hip displacement in children with CP. In 2007 Hagglund et al[37] looked at the most widely clinically accepted thresholds used for AI to categorise 'at risk' hips. The thresholds often adopted in clinical practice are either greater than 27° or greater than 30°. The team observed that AI and RMP were often increased together, however RMP often reached the 'at risk' thresholds first and not all individuals with increased RMP had increased AI. In 2001, Scrutton et al[43] reported that all hips with an AI of over 30° by 30 months went on to have a problem with that hip by 5 years of age. However, they also concluded that AI was a less sensitive measure than RMP in the younger cohort (18 months to 5 years). Contrary to Cooke et al[23] findings, both longitudinal cohort studies from Terjesen's team[32] and Hagglund's team[44] concluded that femoral head displacement preceded acetabular dysplasia and therefore AI should be used as a supplementary measure to RMP and not as a standalone measurement to describe hip dysplasia.

1.8.3. Alternative measurements of proximal femoral geometry

There are many measurements made from 2D planar X-ray which have been proposed to have predictive value for identification of hip dysplasia, RMP and AI are the most widely used clinically. Femoral head shaft angle (HSA), Neck shaft angle (NSA) and Hilgenreiner's epiphyseal angle (HEA) all measure the proximal morphology of the femur. They are all measured from 2D planar X-ray and are variably influenced by anatomical positioning. In hip dysplasia the HSA and NSA are typically increased and HEA typically reduced compared to the typically developing population.

HSA was included by Hermason *et al*[29] in the CPUP equation, as a factor in predicting the likelihood of a child developing hip dysplasia within 5 years.

1.8.4. Measurements of hip dysplasia in 3D

Whilst RMP and AI are clinically useful and relatively simple to measure, neither describe the 3D nature of hip dysplasia. Computed tomography (CT) scans can visualise the complex geometries to better plan intervention. Whilst CT derived measurements are not routinely used, and would be inappropriate for regular monitoring, they have been used in research studies to validate 2D measurements, improve understanding of the natural history of hip displacement and look for predictive factors.

It has been observed that deformation of the acetabulum is often associated with hip dysplasia and dislocation. Chung et al[45] analysed CT scans of 27 children with CP with either displaced or dislocated hips (defined by RMP measurement from planar x-ray). They observed a difference in the location and extent of the deformation to the acetabulum dependent on the severity of the dysplasia. The displaced group was characterised by defects to the posterior wall of the acetabulum, whereas the acetabulum was more globally affected in the dislocated group. When compared to an age matched typically-developing population, both CP groups displayed a significantly reduced acetabular volume with the dislocated group more affected. They also noted that the displaced hips had a shallower acetabulum than those in the control group. All cases with severe displacement had a degree of acetabular deformation. The

authors speculated that for full dislocation there must be a defect to the superolateral acetabular wall and a reduced acetabular volume.

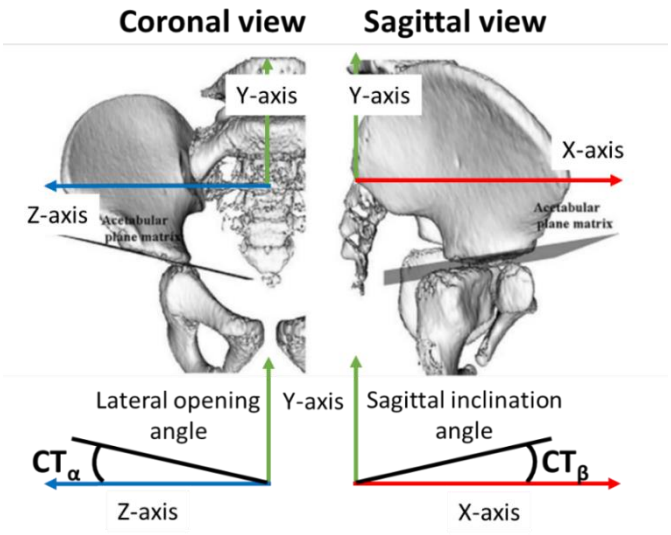


Figure 7: Figure showing the construction of the CTα and CTβ indices (figure adapted from Gose et al[44])

Although it is possible to compute 3D measurements of deformity from CT data, generally 2D parameters are extracted from planar images that are resliced from the 3D volume. The angle between each of the axes of the scan and the acetabulum is measured. The lateral opening angle (CTα) is the projected angle in the coronal plane and the sagittal inclination angle (CTβ) is the projected angle in the sagittal plane (Figure 7). CTα can be considered to be the equivalent of AI in situations where the 3D data is reformatted to align the axis system precisely with the anatomy. Gose *et al*[46] also created a CTMP (Computed Tomography Migration Percentage) index, where centre of rotation of each of the femoral head and acetabulum were used to create a ratio of coverage of the femoral head. Remembering the two elements of hip dysplasia, the team conclude that CTMP is more sensitive to pure femoral head migration in the absence of acetabular dysplasia; but would not be greatly affected in

cases where hip deformity was due predominantly to acetabular dysplasia. CTMP and CTα were found to be positively correlated, whilst CTβ was not found to correlate with CTα or CTMP and thus was not considered to be a useful index.

CT data has also been used to assess acetabular dysplasia. In a retrospective study[45] of patients who had undergone x-rays and 3D CT scans, the correlations between acetabular indices, RMP, GMFCS level and age were investigated. RMP was measured from the X-rays. The CT scans were re-sliced along three planes about the acetabular rim. Similar landmarks to those used to compute the AI were identified, the triradiate cartilage and the acetabular rim. The angle between the edge of the acetabulum rim and the plane formed from triradiate cartilage was measured from each of the slices. The indices were named the anterosuperior index, superolateral index and posterosuperior index (Figure 8[47]). A linear mixed-effects model was used to look at the effect of age, sex, GMFCS level and side of affected hip on the indices of hip dysplasia. Age appeared to have a 'corrective' effect on superolateral and

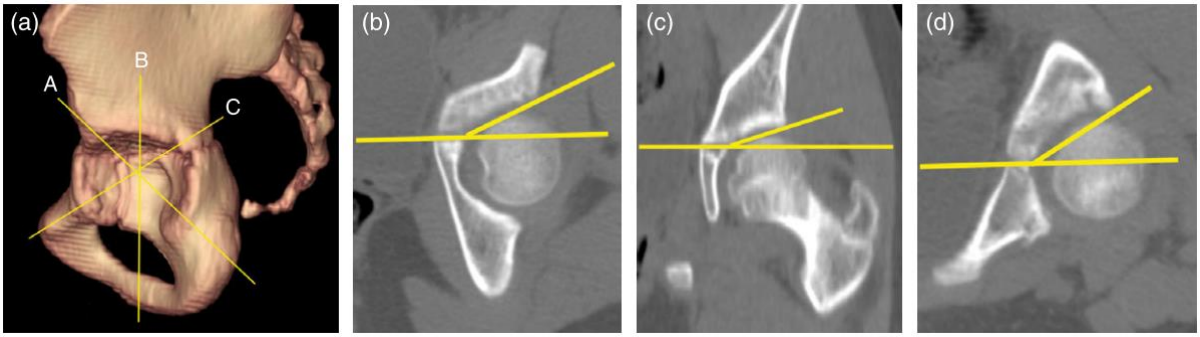


Figure 8: from left to right; a. the reconstructed CT scan showing the 3 cutting planes for the 3 slices; b) anterosuperior plane with marked angle; c) superolateral plane with marked angle; d) posterosuperior plane with marked angle[43].

posterolateral angles which both decreased with age. The anterosuperior index (plane A (image b) in Figure 8) was used to confirm anterior dysplasia of the acetabulum. The superolateral index (plane B (image c) in Figure 8) was used to confirm global dysplasia of the acetabulum and the posterosuperior index (plane C (image d) in Figure 8) for posterior dysplasia. All acetabular indices were associated with GMFCS

level. The team found a significant difference in the posterosuperior index between the control population and the children with CP with functional levels GMFCS I and GMFCS II (p=0.01). They also discovered that GMFCS level was directly associated with acetabular dysplasia regardless of RMP measured from planar X-ray. The authors conclude that in the more physically able individuals, simple 2D radiographic assessment may be insufficient for assessment of acetabular dysplasia. The repeatability of all indices (anterosuperior, superolateral and posterosuperior and RMP) were assessed across 3 orthopaedic surgeons and all showed excellent agreement as measured by intraclass correlation coefficients (ICC).

In contrast, Park et al[48] reported lower inter- and intra- assessor repeatability when computing acetabular indices from CT data. They report inter-operator ICC ranges of 0.7-0.95 for measurements of the same acetabular indices. They found that acetabular indices were dependent on the CT slice chosen to perform the measurements and suggested that this would lead to the mis-classification of many abnormal hips as normal.

1.8.5. Robin and Graham classification system[49]

Due to the lack of homogeneity of the deformities in the group and the errors that have been identified with using a single index, Robin et al[49] aimed to develop a classification system for radiographs to be used clinically to communicate the natural history of hip displacement in children with CP. The system used RMP thresholds along with more descriptive statements to 'grade' the hip (Figure 9). Due to RMP being the most widely clinically used system, they tested this estimated hip grading system against

measured RMP grading system and found excellent agreement. However, the validity of such a comparison is questionable given RMP's inclusion within the scale. The agreement between Shenton's arch, femoral head shape, acetabular shape and pelvic obliquity was above 90% for all and above 96% for all but acetabular shape. Gose *et al*[50] tested the grading system using 3D CT with the aim of validating the system in a younger cohort as they recognised two main limitations of the original study: 1. The use of 2D radiographs to describe the 3D problem and 2. The system was based on data from mature skeletons. Gose et al's[50] study supported the use of the system in the younger cohort with all indices used (CTMP, CTα and NSA from CT) being significantly different between GMFCS levels in children aged 2-7 years. As RMP increased, CTα and NSA from CT increased.

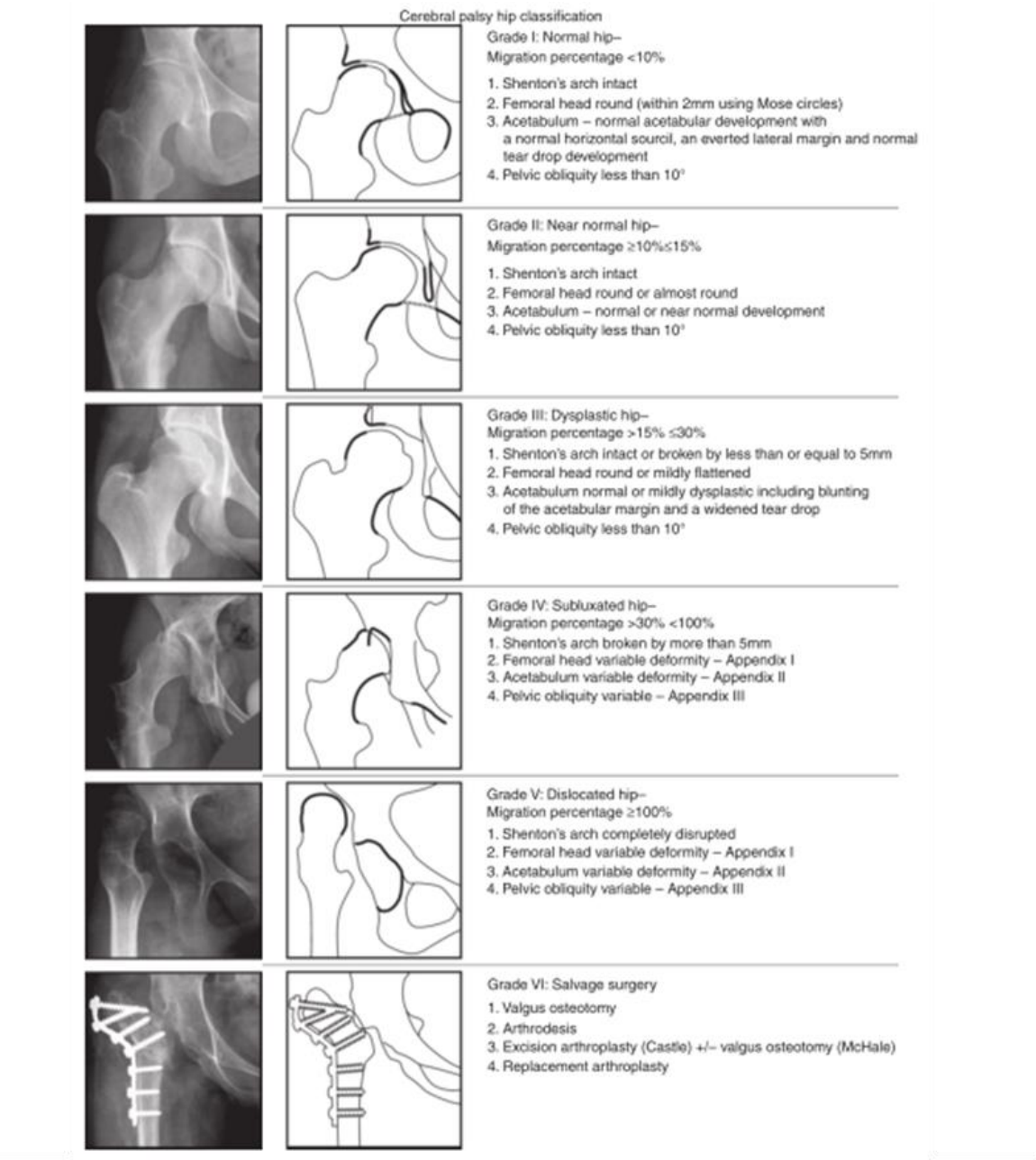


Figure 9: Robin and Graham classification system[47]

1.9. Clinical management of hip dysplasia in cerebral palsy

There are a variety of interventions that have been proposed to prevent and to treat hip dysplasia in CP. The efficacies of these interventions are variably reported, and in some cases remain unproven, perhaps in part due to the lack of consensus on when

different interventions should be considered or are appropriate, Table 1 shows some commonly quoted thresholds. Interventions range from postural management, to botulinum toxin injections, to soft tissue and bony surgeries depending on the severity of the hip dysplasia[21].

Riemer's MP (%)	*Classification*	*Indications*
<33	Stable[27]	No intervention
33 – 89	Displaced or sublexed[27]	Surgical intervention
30-40	At risk	Monitored and conservative measures
>33		Soft tissue releases[32]
>50	Severe displacement or subluxation	Bony procedures[32]
>90	Dislocated[27]	Salvage procedures

Table 1: Common classifications of RMP and associated clinical implication

1.9.1. Postural management

The evidence base for postural support in the management of hip dysplasia is poor. However, a range of devices exist which have been, and are currently used (often in conjunction with other treatments) which have been cited to benefit hip development[51]. These range from standing frames, to kneeblocks on wheelchairs to abduction braces to 24-hour support systems[52–54]. The rationale is to load the hips in a direction that will stimulate bony development of a congruent hip joint. Pountney *et al*[52] conducted a retrospective study with 41 children using postural management before hip subluxation. They categorised the children into three groups depending on the level and intensity of their postural management, ranging from 24-hour management (which included the use of all Chailey Adjustable Postural Support (CAPS) systems i.e. the sleeping, sitting and standing systems), through to just the CAPS, or equivalent seating system. The mean review period was 7 years (1.2 years – 16.9 years) with follow up ages ranging from 3.2 years to 18.4 years. No children in the study could sit

independently. They found a significant difference in the hip status between the groups. With those receiving 24-hour postural support significantly more likely to have stable hips (RMP<33%) than those with just seating or 2 of seating, sleeping or standing support. The hip dysplasia rates were 35% in the group with all CAPS management, 58% in the group with 2 of the 3 systems in place and in the final group with just the seating support it was 89%. The authors conclude that ’24-hour postural management is essential to help direct movement patterns towards ensuring maintenance of muscle length and joint range’ and ‘this retrospective study gives a clear indication that conservative management of hip deformity can be successful if implemented before the development of hip subluxation’. This study lacks the power to be generalised. Functional ability and age are known to be the biggest risk factors in development of hip dysplasia, this study does not control for these factors, further the ability to comply with such intensive postural management is likely to have introduced a selection bias into the study design[55]. The ability to use standing systems may be indicative of a higher functional level, relative to those who did not use them. This alone may explain the difference in the categories. It is also worth noting that the categories were defined by the number of CAPS systems the children had access too, compliance with their use was not documented.

1.9.2. Pharmacological intervention

Boyd *et al*[53,54] conducted a randomised controlled trial of the use of Botulinum toxin A (BTX-A) and a variable hip adduction orthosis (SWASH) to control hip dysplasia. They reported a change in gross motor function, measured by the Gross Motor Function

Measure (GMFM), over a short follow up period. Scores in the GMFM have been linked to risk of hip dysplasia. Their treatment group had BTX-A injections to their adductors and hamstrings at 6 monthly intervals, followed by 6-8 hours daily of bracing using the SWASH orthosis, which holds the child's hips in an abducted position. The children were assessed at baseline and at a 12 month follow up. The control group continued with their standard management, including physiotherapy and postural supports in the form of seating systems. The team reported no treatment effect between the two groups as measured by the GMFM, however they do report a difference between the groups in the surgery rates within the 12 month follow up. 35% of the control group underwent adductor surgery whilst just over 10% of the treatment group required the same soft tissue surgery[54], the authors conclude that the follow up period is too short (12 months) and sample size too small to draw conclusion. Graham *et al*[53] incorporated this cohort into a longer term with a 3 year follow up. After the longer follow up period, with the intervention group receiving the BTX-A injections every 6 months, a very small treatment effect (hip dysplasia rate was reduced by 1.4% per year) of the combined interventions was seen compared to the control group[53]. However, the team reported significant rates of progressive hip dysplasia in the treatment group and concluded that their data does not support the use of this combined treatment in the management of hip dysplasia in children with cerebral palsy.

1.9.3. Surgical Intervention

Surgical interventions can be divided into three categories: preventative, reconstructive and salvage surgeries. Preventative surgeries include soft tissue

releases and muscle lengthening, designed to balance out forces about the hip to encourage better bony development. Reconstructive surgeries, are bony surgeries to either or both of the femoral or pelvic components of the hip designed to reposition the joint for more functional use whilst minimising the chance of future discomfort. Reconstructive surgeries are often also considered to be preventative, i.e they are commonly performed pre-emptively to prevent dislocation. Salvage procedures are very significant bony surgeries where the hip function is compromised, and the reduction of pain and ease of care become the primary motives.

The American academy of cerebral palsy and childhood disability (AACPDM) reviewed the evidence for adductor releases as preventative surgical procedures in the management of hip dysplasia. The theory behind such practices relates to the assumed mechanism of dysplasia, whereby spasticity or increased tone, combined with asymmetrical muscle shortness, namely in the hip flexors and adductors, causes an imbalance of forces about the hip joint, which over time leads to the mal-development of the hip joint[26,27]. Adductor releases aim to reduce the fixed asymmetry in the muscle forces by lengthening the adductors. Early protocols recommended open tenotomies of adductor longus, brevis and gracilis +/- an anterior obturator neurectomy to try to control the increased tone[56]. With time the addition of the neurectomy reduced due to concerns over the permanent weakening and stunting of the adductor muscles' development. Treatment of the contralateral, unaffected, hip also varies with many leaving it untreated but Carr and Gage[57] recommended performing bilateral tenotomies to prevent wind sweeping (the abduction and external rotation of one hip whist the opposite hip is adducted and internally rotated) deformities post-surgery.

In the thirteen studies reviewed by the AACPDM panel in which RMP was used as the index of hip dysplasia, 51% of hips improved post adductor release, 26% of hips

progressed further and 23% of hips remained the same after adductor releases[56]. Six of the thirteen studies reported a significant improvement on passive range of motion post adductor release. No study was large enough to reliably investigate confounding factors, however 11 of the 13 studies analysed age as a factor and 8 reported no difference in outcome related to the age at time of surgery. The other 3 studies reported better outcomes in the younger cohorts. 2 of the studies split the cohorts into those with spastic diplegia and those with spastic quadriplegia. They both reported that those with diplegia had a better response to the adductor surgery. Cottalorda *et al*[58] concluded that adductor tendonectomies were useful in re-positioning the femoral head but not in correcting hip dysplasia. They divided their cohort into 3 groups depending on their RMP at initial presentation, <20% RMP, 20%-40% RMP and >40%RMP. The average follow up time was 6 years, in the low RMP group only one hip was not classified as stable (within 10% of original RMP). In the middle group 48% were stable, 28% were good (a decrease in RMP by greater than 10%) and 24% were bad (an increase in RMP by more than 10%). In the group that started with an RMP of greater than 40%, 35% were stable and the rest the RMP increased by more than 10%.

Reconstructive surgeries are typically performed if soft-tissue surgeries have not prevented further progression or the level of hip dysplasia has exceeded 40% RMP[59]. At this point, without significant change in femoral, and often pelvic, geometry the displacement is likely to continue to progress to dislocation. Varus de-rotation osteotomies (VDRO), +/- shortening of the femur, +/- pelvic osteotomies, +/- soft tissue releases (described above), are performed. VDRO involves cutting the femur at the femoral neck and de-rotating it to correct for excessive femoral anteversion, and re-angle the neck to create a varus neck shaft angle by a varisation. Shortening of the

limb is often combined to reduce the force from hypertonic or short muscles[35]. Where necessary, a pelvic osteotomy is performed simultaneously, with the aim to create a deeper acetabulum to better fit the femoral head and reduce the chances of displacement in the future. Al- Ghadir *et al*[60] reported a 25% revision rate in individuals who received a VDRO alone verses 0% revision rate in those who received VDRO plus San Diego osteotomy. The Dega osteotomy, (first described in Polish[61]), and adaptions of the Dega osteotomy, including the San Diego osteotomy[62] mentioned above, involve the insertion of a bony wedge superior to the acetabulum (Figure 10[35]). This increases the curvature and global lateral coverage of the acetabulum, including posterior-superiorly which is key for individuals who are seated most of the time. Such procedures have been shown to have excellent long-term stability[63–68]. There are differing opinions on the timings of these procedures, with the risk of bone remodelling in the young re-deforming the proximal femur[69], whilst for the skeletally mature the triradiate cartilage has closed which is a contraindication for both the Dega and San Diego osteotomies which rely on acetabular hinging on the open triradiate cartilage. However Murar *et al*[67] found no statistically significant differences between the outcome of San Diego osteotomies in the those with closed and open triradiate cartilage challenging this view. There is also differing practice around treatment of the contralateral hip for levelling the pelvis, however this is still widely disputed[59].

In cases where the hip is severely displaced , or even dislocated, an open reduction may be required alongside femoral and/or pelvic osteotomies. An open reduction involves opening out the capsule to allow the femoral segment to sit within the socket to form a more congruent joint[70].

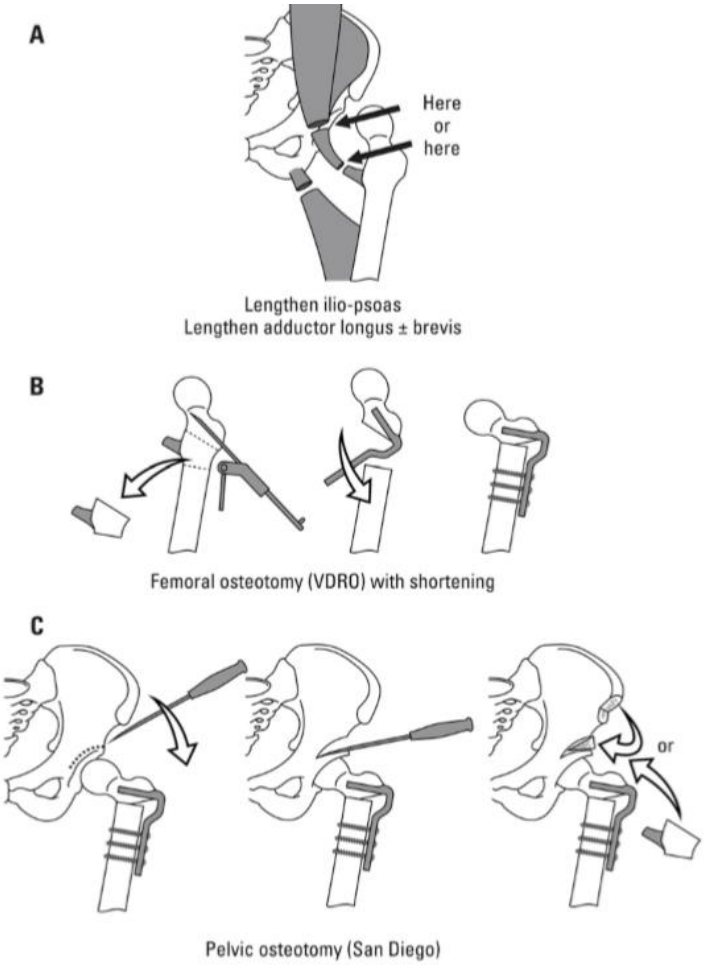


Figure 10: Schematic of different preventative and reconstructive surgeries[34]

Salvage surgeries are considered as a last resort and the hip is deemed irreducible due to degeneration or severe deformity, secondary to the displacement. There are several surgical options, outlined by Shore and Graham[35], however with all options the goals are limited to improving comfort and perineal hygiene. In most cases hip function is severely impaired. The move towards routine monitoring of hip dysplasia in this population has seen a dramatic decrease in dislocations, resulting in fewer salvage procedures[24].

Whilst there is little consensus regarding the specifics in the management of hip dysplasia there is agreement that earlier identification is important. Hip surveillance programmes have been established world wide in an attempt to identify those at risk of developing hip dysplasia and facilitate the monitoring of these individuals with the aim to ensure optimal timing for intervention.

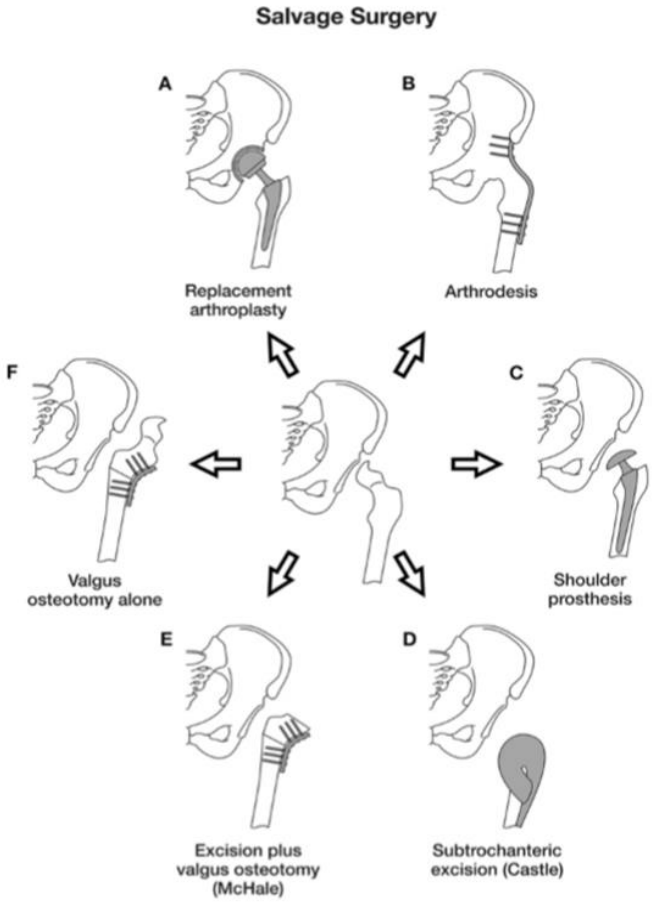


Figure 11: Schematic of salvage surgeries[34]

1.10. Hip surveillance programmes

Hip surveillance programmes have been reported to dramatically reduce the incidence of hip dislocation[24,44,71]. The first large-scale population-based hip surveillance programme started in Sweden in the 1990s. Under this programme, hip X-rays were performed once-a-year until 8 years of age. Twenty years after the programme began the incidence of hip dislocation had decreased from 8%, prior to standardised monitoring, to 0%[24]. Preventative surgery, either soft tissue releases or bony surgery, was performed on 13% of children in the programme[24]. The success of the Swedish initiative has been replicated in similar programmes across the world[71], although the timing of assessments and degree of displacement indicating intervention varies between programmes. Most surveillance programmes have stratified individuals by functional level. Individuals receive an annual X-ray and a clinical assessment of their hip range of motion. Initial assessment is at approximately 2 years of age followed by annual assessment of non-ambulant individuals until skeletal maturity. Although the passive range of motion of the hips is often recorded, Hagglund *et al.*[30] concluded it was a 'poor indicator of hips at risk'. Hermanson *et al*.[29] developed a predictive scale, designed to predict the likelihood for an individual of developing hip displacement in the next 5 years. Passive range of motion of the hip was not included in their predictive algorithm, despite the data being collected.

The success of surveillance programmes has primarily been quantified by the reduction in the incidence of hip dislocation. But this is not the only consideration in the management of the hip in children with cerebral palsy. Function, pain and quality of life, as well as other environmental and personal factors, are also considered but rarely evaluated as an outcome of hip surveillance programmes. Disability and pain are associated with hip displacement well before the end-point of dislocation[72].

Wawrazta *et al*[72] looked at a cohort of adolescents and young adults with cerebral palsy and hip displacement, some of whom had been under hip surveillance programmes and some who had not. They reported that on average pain was more severe and more frequent in the group who had not been under surveillance, presumably due to the lack of regular contact with clinical services.

Optimisation of monitoring intervals and intervention thresholds are dependent on expected progression rates of hip displacement and the accuracy and precision of the measurement.

1.11. The influence of skeletal changes on measurement of RMP from X-ray images

As discussed previously, ossification of the hip complex is a dynamic process which is not completed until skeletal maturity. As cartilage is not well-differentiated from other soft-tissue on X-ray, there may be an overestimate of hip displacement from X-rays in the young hip. As cartilaginous tissue is replaced by bone, this error is likely to reduce. Therefore age, or more specifically level of skeletal maturity, will influence the accuracy of the measurement of hip displacement from a 2D planar X-ray. In children with CP, and particularly those with greater delays in motor development, the process of ossification may be altered. These infants do not load their joints as early, or as often, and may have altered muscle development in comparison to their typically developing peers[73]. Hermanson *et al.*[29] found age to be a predictive factor in the risk of an individual developing hip displacement. However, there is a lack of empirical data necessary to include an 'age factor' when defining standardised thresholds for

intervention, but it is likely included in the tacit understanding of experienced clinicians when making intervention decisions.

In 2007, Hagglund *et al.*[37] described a group of children who had initially presented with increased hip displacement (indicating intervention was required) but subsequently presented with less displacement, without receiving any, or minimal, intervention in the interim. “Spontaneous correction” as observed by Hagglund and his colleagues could be explained by; a genuine improvement in hip development, increasing ossification of the hip in the developing child leading to systematic changes in the appearance of planar X-rays, regression to the mean resulting in a reduction of the measured RMP, or purely a product of measurement error. To our knowledge the effects of ossification on measurement errors in RMP and surgical prescription have never been quantified.

1.12. Position and projection errors

Ossification of the hip in early childhood may give rise to systematic errors in RMP but, even in a completely ossified skeleton, measurements of 3D bony anatomy from planar X-ray are subject to potentially large errors. The source of errors is perhaps inherent in the 2D X-ray used to quantify RMP and other measures of proximal femoral morphology. The hip radiograph is a 2D projection of a 3D geometry. The content of the projected image changes depending on the orientation of the body to the plane of the image (projection errors). As well as the orientation of the whole body, the relative orientation of body segments influences the X-ray image content (position errors). For example, an internally-rotated and adducted hip will appear to have an increased RMP in comparison to a neutral hip and externally-rotated abducted hips will appear to have

a reduced RMP. These problems affect the precision and accuracy of the RMP measurement. Windswept hip deformity, where one hip becomes fixed in an externally rotated abducted position and the other in an internally rotated adducted position, has a relatively high incidence (approximately 12%[74]) in children with CP, and is particularly prevalent in non-ambulant children.

Lateral displacement of the femoral head relative to the acetabulum is often considered to be the primary direction of pathological hip displacement, however, any component of the displacement perpendicular to the plane of the X-ray image is not quantified. Brunner *et al*[25] used computed tomography to investigate 24 hips that had fully dislocated. In all cases a clear groove was present in the roof of the acetabulum, indicating the direction of displacement of the femoral head. The mean direction of the channel was lateral, but the range of directions of the channels extended from 33° anteriorly to 70° posteriorly. An X-ray measurement in the coronal plane, such as RMP, would underestimate any displacement that is not within the plane of the image.

1.13. Repeatability of measurements from planar x-ray

There are no studies on the repeatability of the complete process of acquiring an X-ray image of the hip in individuals with CP, including repositioning of the patient and retaking of the X-ray at an appropriate time interval. Rather, estimates of the repeatability of RMP have been determined from repeated measurements on the same group of X-ray images between and within assessors. The lack of consideration of error from repeated image acquisition and the choice of summary statistic may have led to the conclusion that X-ray imaging of the hip is a reproducible method and it has

become the clinical standard technique in this domain. In only one study has the repeatability of X-ray image acquisition of the hip been investigated. Cliffe *et al*[75] investigated the repeatability of their positioning protocol by performing repeated X-rays spaced by at least one hour. They reported that it was not possible to confidently detect a change in RMP of less than 11.5% for repeated measurements on the same subject. However, because the interval between image acquisitions was so short, and they were performed under research conditions, it is likely that this figure is an underestimate of the measurement errors encountered in clinical practice.

Authors	ICC	No of hips	Subject type	Study details	MDD
Craven et al[76]	0.93	228 hips	CP all GMFCS levels	Prospective study Aged 18 months to 5 years Repeated measures from same X-rays 2 raters at least 2 weeks apart	10.8%
Kinch et al[77]	-	20 hips (40 including replication)	CP all GMFCS levels	Retrospective study Repeated measures from same X-rays 5 raters (interrater data taken)	11.0%
Cliffe et al[75]	0.96	40 hips (repeated images)	Bilateral CP GMFCS levels IV and V	Aged 30 months to 10 years Repeated images spaced by an hour 2 raters at least 3 months apart	10.3%
Parrot et al[78]	0.91	20 right hips	Bilateral CP all GMFCS levels	20 X-ray selected from 110 X-rays, insuring images across a range of GMFCS levels and RMP levels. 5 raters (interrater data taken)	11.5%

Table 2: Published repeatability of RMP studies with corresponding minimal detectable differences (MDD)

The reproducibility of measurement is often expressed by the intra-class coefficient (ICC)[75,76,78], the ratio of inter-subject variance to the total variance (including the variance owing to measurement error). Often, clinical measurements are categorised as highly reproducible according to the Landis and Koch[79] criteria because the inter-subject variation is large in the sample population. In studies of reproducibility of RMP

in children with CP, the range of RMPs in the samples are typically from around 0 to 60%[78]. This figure is considerably larger than the range of migration percentages over which decisions to intervene are made (from 30% to 50%). Thus, the ICCs calculated in these studies over-estimate the repeatability, and therefore, also over-estimate the ability for RMP to detect important changes in the displacement of the hip and thus its clinical utility. Alternatively, reproducibility can be quantified by the minimal detectable difference (MDD); the smallest change in a measurement that can be confidently considered to be a true change. Table 2 shows the calculated MDD for RMP reproducibility studies in the literature where sufficient data was published. The acceptable level of measurement uncertainly in a clinical index is, in part, driven by the expected rate of change of the pathology being measured. Terjesen published natural history data on hip dysplasia progression rates stratified by GMFCS level and sub-type of cerebral palsy[32]. He quantified hip dysplasia using RMP. Terjesen[32] reported mean annual progression rates varying from 0.2% (s.d. 3.7%) for individuals with GMFCS level I to 9.5% (s.d. 9.4%) for individuals with GMFCS level V. Park *et al*[36] reported lower progression rates of 0.3%, 1.9% and 6.2% for GMFCS levels III to V respectively. There was a significant difference between each of the groups, supporting GMFCS level being a significant risk factor. The mean reported annual rates of progression are in most cases lower that the reported repeatability of RMP. Despite high ICCs (Table 2) the MDDs are large, particularly considering the expected annual progression for some of these individuals. When defining thresholds for clinical utility, it is important to consider the properties of the measurement alongside the natural history of the pathology.

1.14. Is there a risk of unnecessary intervention in hip surveillance programmes?

The success of hip surveillance programmes in detecting and managing hip displacement is undisputed[24,71]. However, the potential deficiencies of these programmes must be appreciated and considered. RMP, which is nearly universally used, is a simple measurement of a complex three-dimensional and articulating joint as it undergoes dynamic development. It is conceivable that it gives rise to occasional misclassifications, i.e. falsely classifying a hip as at risk when it is not, or falsely classifying the hip as normal when it is partially displaced.

1.14.1. The consequences of misdiagnosing true hip displacement

A false negative may result in a displaced hip not being directed towards the most appropriate intervention after assessment. However, if a hip with significant displacement is missed at the first assessment, it is likely, at the next assessment (typically a year later), it will be detected. 'Missing' a displaced hip may result in further deterioration of hip function, increased pain and further displacement in the interval between assessments but is unlikely, given average progression rates, to result in dislocation.

1.14.2. The consequences of falsely identifying hip displacement

Incorrect classification of a child as having hip displacement (a false positive) may have greater clinical significance. Assessing the likelihood of a false positive result is challenging and unquantified in the literature as these children will have been indicated for, and may have received, surgery. Measurement error may lead to a child

undergoing unnecessary interventions, exposing them to unnecessary risks. These could include surgical and anaesthetic risks, unnecessary pain or missed educational, recreational and social opportunities. The design of hip surveillance programmes is such that over the period of surveillance, the programme may be sensitive to detect hip displacement, but may lack specificity, potentially leading to unnecessary treatment.

1.14.3. Implications for radiographic surveillance programme design

Typically, hip surveillance programmes recommend annual review with X-ray and measurement of RMP to assess stability of the hip. However, we know from natural history data[32,36] that progression of hip displacement by greater than 10% RMP in a year is not common, particularly for children who are older and have a greater level of mobility. Yet RMP is not precise enough to detect a true change of less than 10% RMP with a confidence of greater than 95%[75–78]. There is some suggestion that bi-annual screening may be appropriate for some children[39] but this may lead to a greater risk of a false positive findings as mean progression of hip displacement will be smaller over a shorter interval and therefore the ratio of measurement error to progression will be greater. Confidence in the reliability of a measurement could be increased by taking repeated measures and averaging the results at the same time point.

Given the acquisition of these X-rays exposes the very young to ionising radiation, consideration of the risk of multiple doses due to serial acquisitions, as well as the rate of progression of the pathology, needs to be included when defining surveillance intervals.

It is common in health screening programmes to use a test that is very sensitive to the underlying pathology. These tests tend have poor specificity, that is they tend to have high false positive rates. Individuals identified as positive by the screening test should then undergo further tests with higher specificity to confirm the diagnosis. These additional tests may carry a greater risk of discomfort or morbidity than the initial test but are preferable to unnecessary intervention[80]. Current hip surveillance programmes may behave similarly to other health screening programmes with high sensitivity and poor specificity due to the poor measurement properties of the X-ray measurements. Hip surveillance programmes may therefore benefit from the addition of a more specific test in those indicated for intervention by planar X-ray imaging, to confirm a positive diagnosis before any significant interventions are performed.

Three-dimensional imaging modalities are likely to provide a good solution, however currently, despite volumetric image acquisition, analysis is still usually from 2D slices within the captured volume[45,48] rather than re-slicing or making a truly 3D measurement. More work is needed to develop and validate true 3D parameters which better characterise hip displacement in this population.

1.15. Limitations

Repeatability studies have their limitations, firstly they are primarily retrospective, meaning that the true effects of position cannot be assessed. Secondly when analysing and making measurements under study conditions it is likely that specific training, equipment and additional care is taken, which does not simulate real world repeatability. Prospective repeatability studies will have standardised protocols for

patient positioning which will be better controlled and adhered too then in the real world. Finally, studies are normally conducted at a single site, which does not account for variability between centres.

Most studies identified for this review utilised retrospective data. The use of retrospective data can limit the design of studies in several ways. Firstly, the limited availability of data may limit the sample size, resulting in an inadequately powered study; or, study designs may be manipulated to ensure sufficient data to conclude and extrapolate results. The latter is likely to contribute to the use of migration percentage thresholds for characterising cohorts of separating groups, as opposed to more detailed measures of hip morphology. This may be inappropriate if the research questions focus on treatment or intervention for hip dysplasia and measures its outcome by the same measurement (RMP) used to characterise the study participants. It is possible that these studies become susceptible to confounding errors introduced at the separation of the study groups. Further, using an inadequate proxy for hip status may mask genuine changes in hip morphology as a result of the treatment or intervention under investigation.

1.16. Summary

In summary, hip surveillance programmes for children with CP have been shown to improve the identification and timely treatment of hip displacement[24,71]. Their efficacy has been demonstrated by the dramatic reduction in incidence of hip dislocations, but more investigation is needed to understand the likelihood of measurement uncertainty impacting treatment decisions. This was investigated further in the present work

(Chapter 3) using simulation. Surveillance programmes are designed to be highly sensitive however, the specificity of the programmes remains unreported and unquantifiable by clinical study, potentially concealing a hidden group of children who receive unnecessary intervention[37]. Surveillance programmes may be improved by further imaging, particularly for those diagnosed with clinically-significant hip dysplasia by planar X-ray imaging, to reduce the risk of unnecessary intervention. However, considerable work is needed to develop, validate and assess the accuracy and feasibility of true 3D parameters for quantification of hip displacement in this population.

2. Technical introduction

2.1. Overview

The previous chapter provided a motive for the thesis, it details a clinical issue and the current chosen management techniques. This chapter is a technical introduction or background chapter, it provides introductory material on concepts and techniques that are used in the studies presented in the latter chapters. There are three main sections, the first two, imaging and surveillance programmes, are written to provide a broader understanding of the techniques used in the studies described and the theoretical basis for them, addressing the underlying principles for current and potential future practices. The third section, validation, provides the background for the methodologies used in this thesis to address some of these problems.

2.2. Imaging

2.2.1. X-ray and related imaging modalities

The standard imaging modality used to assess hip dysplasia is X-ray. Planar radiographs are routinely taken to visualise the positioning of the bony segments of the hip. X-rays are electromagnetic waves, they have a shorter wave length and higher frequency than visible light, the combination of these properties allow them to pass through many media. Different media absorb the waves to different extents. X-ray machines utilise this principle to generate images of anatomical structures.

X-rays are produced in X-ray tubes. The tubes consist of an anode and a cathode inside a vacuum. A voltage difference is generated across the tube to generate electrical current flow. Using a high voltage power source the cathode is heated and emits electrons which are accelerated towards the positive anode. The interaction between the electrons and the tungsten nuclei results in the emission of energy in the form of X-rays which are then directed towards the patient.

There are two mechanisms that generate X-rays, characteristic X-ray generation which makes up approximately 20% of the X-rays within a beam, and Braking (or Bremsstrahlung) X-ray generation. In characteristic generation, high energy electrons are accelerated towards tungsten molecules, the electrons collide with electrons in the inner shell of the tungsten atoms, displacing them. An outer layer electron is then promoted to the inner shell, resulting in emission of energy, which is an X-ray photon. In Braking X-ray generation, the X-ray is emitted as a result of the electron decelerating as it approaches the nucleus, this causes the electron path to be defected and energy to be emitted. Approximately 80% of X-rays within a beam are generated in this way.

The properties of the X-ray beam are altered by changing the applied voltage or the anode material and by installing aluminium filters of various thicknesses. The radiation dose the patient receives, is controlled by exposure time and adjusting the current flow[81].

The X-ray beam is directed to pass through the body, the different tissue types absorb energy from the X-rays to different extents, depending on the radiological density of the tissues they pass through. The detector measures the quantity of waves as they reach the detector and display a grey-scale image. It is important to note that the resulting image is a result of cumulative absorption along the ray trajectory.

Radiological density of tissues is dependent on both the density of the tissue and the atomic number. Bone, for example, has a significant calcium content, which has a high atomic number and therefore bone absorbs X-rays well, creating a 'shadow' on the resultant radiograph. The difference in radiological density of tissues is routinely used in medicine to identify abnormalities, both bony but also in soft tissues – for example identification of tumours in mammography. Dependent on the purpose of the imagining, the energy of the X-ray can be changed to identify different features. For example, in bone mineral density imaging (DEXA scans), two different energies are used. In the same way as in a standard X-ray, a detector is used to measure the rays that have not been absorbed. The dual energy X-rays allow for an estimate of soft-tissue absorption, using a lower energy X-ray, and the higher energy X-rays can, in part penetrate the bone. The absorptions of the different tissues can be estimated facilitating the estimate of density in different bones and comparisons to normal data sets. These measurements are used to diagnose conditions such as osteoporosis[82].

Computed tomography (CT) also uses X-ray to visualise structures non-invasively. A narrow beam of X-rays is directed towards the patient, the beam is rotated around the body. The acquired signals are processed, and cross-sectional slices are captured. The patient is positioned on a bed that is slid through the rotating X-ray beam. This allows multiple cross-sectional images, or slices to be acquired and 'stacked' together to create an image volume. This volume data can be manipulated and re-sliced to create 2D images in any orientation. Like simple planar X-rays, CT scans can be used to look at some soft tissues, for example they are used to detect trauma, tumours or bleeds in the brain, however they are often used for visualising bony structures in 3D.

The frequency and wavelength of X-rays mean that they can cause adverse effects in human tissues, either by causing DNA damage or alterations to intra cellular processes. For this reason, exposure to X-rays is controlled and limited. We are all exposed to a background radiation dose, which varies significantly and depends on geographic location, on average approximately 2.2 mSv per year. A sievert, or Sv, is a unit of radiation dose used to quantify the biological effect of radiation, 1 Sv is equivalent to 1 joule of radiation energy in 1 kilogram of human tissue). Typically, diagnostic radiation doses are in the range of 0.02 mSv, equivalent to a few days of background dose for a planar chest X-ray, through to 10 mSv equivalent to approximately 5 years of background dose for a CT abdomen.

EOS® is a relatively new, bi-planar X-ray technology which uses significantly less radiation than CT scans and traditional X-rays. Frontal and lateral radiographs are taken simultaneously, and algorithms and data processing techniques are used to create 3D images. Despite this large advance in X-ray technology, EOS® is not yet routinely used in most clinical services. Adoption is likely to be limited by the requirement to have a highly skilled radiographer to ensure that the data is processed

appropriately to create the high-resolution images and 3D renders from the low dose X-rays and the high cost of the technology. Regardless, EOS® has been shown to have great potential in the research domain, and could in time prove invaluable clinically, possibly even in the routine monitoring of hip dysplasia[83].

2.2.2. Magnetic resonance imaging

Magnetic resonance imaging (MRI) relies on the properties of hydrogen nuclei, which are ubiquitous including in the human body, to create detailed 3D images. A hydrogen nucleus is a single proton, which has a quantum mechanical property called spin. In the absence of an external magnetic field, the hydrogen atoms in the body spin with their axes of rotation randomly orientated. An MRI scanner consists of a large magnet, when a person is placed inside the magnet the axis of rotation of the hydrogen nuclei align parallel or anti-parallel with the external magnetic field.

Within an external magnetic field, hydrogen nuclei exhibit precession where the spin axes of the nuclei rotate around the direction of the magnetic field. The frequency of this rotation, the Lamor frequency, is proportional to the strength of the applied magnetic field. There are always more nuclei that align with the external magnetic field than anti parallel, this results in a net magnetisation aligning with the external magnetic field. This magnetisation from the hydrogen alignment is too small to detect directly, instead a radio frequency (RF) wave is pulsed into the body at the same frequency as the Lamor frequency, and this causes the protons to resonate and momentarily deflects the axis of magnetisation into the transverse plane. After the cessation of the RF burst the magnetisation continues to rotate, or precess about the direction of the main magnetic field. This rotating magnetisation induces a current in an antenna or detector placed near the body part being imaged, known as a receiver radio-frequency

coil. The amplitude of the received RF wave is proportional to the resultant magnetisation from the precessing hydrogen protons.

To create an image it is necessary to determine how much magnetisation, or how many hydrogen protons, there are in different parts of the body to create a spatial image. This is done by applying a gradient to the external magnetic field. The Lamor frequency is directly proportional to the strength of the external field and thus the RF-wave that is generated by deflecting the precessing atoms can be altered by varying this strength of the applied magnetic field. By increasing the strength of the magnetic field on one side of the body compared to the other, the Lamor frequency of the atoms on one side will higher than on the other. This results in a different frequency of RF-wave that will be detected by the receiving coils on one side to the other. By looking at the frequency of the received RF-wave it is possible to determine where the signal is coming from. By repeating the acquisition, applying different gradients in different directions it is possible to determine the spatial distribution of the hydrogen protons within a body part.

Finally, the contrast in an image comes from the behaviours of the different tissue structures, which contain the hydrogen protons, in a strong magnetic field. As the spins are precessing around in the transverse plane they also begin to re-align with the external magnetic field. As they do this the strength of the transverse, and detectable, magnetic field gradually reduces. This is known as relaxation. The time taken for the precessing spins to align with the external magnetic field is known as the longitudinal relaxation time or T¬1 relaxation time. Different tissues have different relaxation times. If the spins are excited again, by another RF pulse, before they have fully recovered their alignment with the external magnetic field then there will be less signal produced from the next excitation. By changing the time period between exciting

the spins with RF pulses and letting them relax again it is possible to differentiate between different tissues by the amplitudes of the detected transverse magnetisation. In T1 weighted images greater signals are detected from tissues with shorter longitudinal relaxation time, which will appear brighter on the image.

There are two types of relaxation, the other is transverse relaxation. When the applied RF pulse causes the magnetisation to be tipped into the transverse plane and then begin precessing in the transverse plane, there is exchange of energy between adjacent spins which cause an exponential decay of the transverse magnetisation. The rate at which the decay occurs is different for different tissue types, the time constant associated with this decay is called the transverse relaxation time or T2. By acquiring an image at a specific time after the RF pulse it is possible to characterise the tissues based on the T2 relaxation time, areas with a shorter T2 time will have less signal than those with a longer T2 time. This is known as T2 weighted imaging.

The different make up of tissues mean that the relaxation properties vary. This principle is used to emphasise certain properties in the tissues. Pulse sequences have been developed to leverage these differences in different tissue types to highlight different features. For example, a fat suppression sequence allows for fat to be removed from the image, and only signals from abnormalities within the fatty tissues are detected.

MRI relies on nuclei possessing the ability to spin, in theory any nucleus that has this property could be imaged using this methodology. The ability for a nucleus to process spin is dependent on an odd number of either protons or neutrons. Hydrogen is the most commonly targeted nucleus due to its abundance, particularly in soft tissue in the

form of water. It is for this reason that MRI is usually used for soft tissue imaging although it is possible to get bone-optimised MRI imaging.

2.2.3. Ultrasound

Ultrasound is a high frequency sound wave, with the frequency determined by the function required. Piezo-electric crystals in the transducer convert electrical energy to mechanical oscillations (sound wave) and then back to electrical energy as the transducer received the reflected waves. At each tissue boundary part of the ultrasound wave is reflected. The amplitude and angle of the reflection depends on the acoustics difference between the two media. When there is a significant difference in impedance at a boundary, for example between soft tissue and bone, there is a near complete reflection and any deeper structures are shadowed. The direction of the beam changes (refraction) relative to the impedance difference at the boundary, this can cause artefacts.

There are many interacting variables that affect the behaviour of ultrasound. The speed (c) at which an ultrasound wave propagates through a medium is affected by both the stiffness (κ) and the density (ρ) of that medium. Propagation speed increases if stiffness increases or density decreases (Equation 2).

$$c = \left(\frac{\kappa}{\rho}\right)^{\frac{1}{2}}$$

Equation 2: Relationship between propagation speed (c), stiffness (κ) and density (ρ)

The acoustic impedance (z) is the resistance experienced by the sound wave being transmitted through a medium which is directly proportional to the density (ρ) of the medium and the propagation speed (c) (Equation 3).

$$z = \rho c$$

Equation 3: Relationship between acoustic impedance (z), density (ρ) and propagation speed (c)

Attenuation coefficient is the parameter used to estimate the reduction in amplitude of the soundwave as a function of the frequency of the wave, attenuation coefficient increases as frequency increases. The direct consequence of this is that penetration depth reduces as frequency increases. Increasing frequency is advantageous for resolution, both axial and lateral resolution.

There are several different types or modes of ultrasound. The most basic being A-mode where the transducer sends a single pulse. The pulse propagates through the media, being partially reflected at boundaries between media with different acoustic properties. The probe receives the reflected sounds waves and converts to an electrical signal. The resultant 'image' is a series of peaks of different amplitudes at different times. The distance between the peaks can be used to calculate the depth of different boundaries. However, there is no further spatial information. B-mode ultrasound is similar to A-mode however instead of a single pulse, a series of piezo electric crystals send out pulses asynchronously. The amplitude of the reflections are turned into a greyscale value with the intensity of the reflection represented by the brightness (hence B mode) which the ultrasound machine processes and displays as a 2D image of depth and distance in the plane of the transducer. There are three methods of 3D ultrasound, all derived from 2D B-mode ultrasound.

2.2.3.1. 3D ultrasound

There are three different types of 3D ultrasound commonly used, each constructs 3D volume data by capturing a series of 2D images and interpolating between the images. Each of the three techniques is appropriate in different situations.

3D freehand ultrasound uses a standard 2D linear probe and a system for monitoring the position of the probe in space, magnetic tracking or commonly motion capture tracking. The system captures a series of standard 2D ultrasound images at known positions in a volume. In its simplest form measurements between images at known distances can be taken. Software solutions have been developed to 'stitch' images together by interpolating between data points in consecutive images to create volume data, allowing for true 3D measurements to be made.

3D sector scanning is a type of 3D probe, an electric motor sweeps the ultrasound transducer through an arc within the probe. A series of high speed 2D B-mode images are taken at different positions in the sweep allowing for images to be stitched together to form the 3D volume. The size of the field of view within the capture volume can be varied by varying the depth of the scan and using gel pads to increase the gap between the probe and the skin surface.

Finally, 3D array probes have a 2D array of piezo-electric elements that simultaneously emit and receive the sound waves to construct the 3D volume. These probes are very similar to multiple linear probes stuck in a line. There are advantages and disadvantages of each of the technologies which lend them to different utilities. 3D freehand ultrasound is capable of capturing large volumes of data, for example the whole length of a muscle[84], whereas 3D array probes and sector scanners can only capture small volumes. External software solutions that can read and synchronise

information from multiple inputs are required for effective us of 3D freehand ultrasound, with a lag between image capture and display of the capture volume. In contrast 3D array probes capture and display data in a similar time frame to standard 2D ultrasound images, with a sector scanner taking up to a few seconds to capture and display the data.

3D ultrasound has proven very useful in soft tissue imaging[84–86], but there are few studies of proximal femoral or hip geometry using the technique. Passmore et al[87] used freehand 3D ultrasound to measure femoral neck anteversion angle comparing results to those obtained from MRI. There was an average difference of 1.8° between the imaging modalities across the 10 subjects. The 3D ultrasound was found to have repeatability coefficient of 3.7° with was comparable to that of MRI, which was reported as 3.1°. The Pearson correlation coefficient was 0.94.

2.2.3.2. Ultrasound imaging of the hip

The use of ultrasound to evaluate the hip in young infants has transformed the screening of developmental dysplasia of the hip (DDH). DDH describes “a range of hip abnormalities affecting the newborn in which the femoral head and acetabulum are in improper alignment or grow abnormally or both[88]”. Clinical guidelines suggest that every infant should be screened for DDH[89]. In the UK, all hips are screened by a clinical examination shortly after birth. For unstable hips, as classified by physical examination, or infants who are considered high risk (female with family history or breech position in the womb), ultrasound imaging is recommended to confirm hip dysplasia. There is an opportunity in the infant to use ultrasound imaging to visualise

the hip and acetabulum because the hip is largely cartilaginous in the first few months of life. As the hip ossifies, it becomes impossible to get the same clear images of the joint. However, it is still possible to visualise significant anatomical landmarks and make measurements of hip geometry which may have diagnostic value. There have been only a small number of studies that have looked to use 2D ultrasound imaging of the hip in children with cerebral palsy. Smigovec *et al.*[90] visualised the hip in children with severe CP (GMFCS IV, V) using 2D ultrasound. They used the scanning technique first described by Terjesen et al[91]. Smigovec et al[90] produced encouraging results, with greater than 90% sensitivity and specificity they discriminated between measurements above or below a cut off RMP of 33%[90]. Prior to their work, Tegnander and Terjesen[91,92] investigated the feasibility and reliability of using ultrasound to assess and monitor the fully ossified hip in children above 2 years. Initially they looked at 'normal hips' i.e. children with no previous hip pathology, normal range of motion (ROM) at both hips and normal anatomic structures around the hip as imaged by ultrasound. All children underwent both anterior and lateral ultrasound scans. Lateral head distance (LHD), defined as "the distance from the lateral tangent of the ossification centre of the femoral head to the lateral bony acetabular rim[91]", and lateral cartilage distance (LCD), defined as "the distance from the lateral tangent of the cartilaginous femoral head to the lateral bony acetabular rim[91]", were measured from the lateral scans – although the team deemed LCD did not add any further useful information. LHD was considered a measure of coverage of the femoral head by the acetabulum. For the anterior scan, anterior head distance (AHD) and anterior cartilage distance (ACD) were measured. AHD measures the anterior coverage of the femoral head. The team concluded that the required relevant bony landmarks could be visualised by ultrasound. They proposed normal limits for LHD measurements of 4 mm

for 2-3 yrs, 5 mm for 4-7 yrs, 6 mm for 8-11 yrs, 7 mm for 11+ yrs. Applying these limits, patients with LHD lower than these limits were categorised as not having hip dysplasia and patients above these thresholds were sent for radiographs.

2D ultrasound has proven to be useful but has limitations related to inter-operator variance[93]. This may be related to the visualisation of a complex 3D geometry with a 2D technique. In contrast, 3D ultrasound is proving to be an accurate and reliable tool in the morphological evaluation of the musculoskeletal system[86] and may be relevant to investigation of hip morphology.

2.2. Surveillance programmes

2.2.1. Principles of screening and surveillance programmes

Screening is used to identify people who have developed a condition before signs or symptoms. Screening programmes can be targeted at specific high-risk groups or can be population wide. The aim of screening programmes is to identify individuals at high risk of developing the disease and prevent or reduce the risk of the disease for that individual. Surveillance programmes systematically collect, analyse and report data from a population known to have the disease. Surveillance data should be used to understand disease trends and predict future trends. Fundamental to the success of both screening and surveillance programmes is an understanding of the natural history of the condition. Surveillance can be divided into two categories, passive and active. In passive surveillance the case has already been identified, data relating to the case is collected and reported to a central system to ensure an understanding of the

population data, often at regular intervals to allow for trends in the population to be identified. Active surveillance is the active search for all the ‘cases’ of the condition within a population, using routine perspective data collection, active surveillance can involve the process of screening[94,95].

A screening programme can be broken down into 3 components, the test, the disease and the preventative action. The perfect screening scenario would be one where a test exists that perfectly discriminates between those who will and will not develop the condition/disease. The disease/condition would be predictable where the outcome without intervention is known and the intervention is a perfect treatment that prevents the disease/condition from developing. In reality, the test rarely has a binary result, instead thresholds are often defined to maximise true positive results and minimise false negative results even at the cost of increased false positive results. Whilst at a population level the natural history of the disease/condition may be known, the development of the disease/condition within a specific individual may be hard to predict. Preventative interventions have a treatment effect, there is rarely a perfect solution for all, and the outcome of a treatment for an individual is often hard to define. Considering these interactions, the performance of a screening programme across a population is assessed by the sensitivity and specificity of the programme. Simply there are four outcomes that can be achieved, a positive test result and positive for the disease (true positive (TP)), a positive test result without disease (false positive (FP)), a negative test result without the disease (true negative (TN)) and negative test result with the disease (false negative (FN)). The ratios of these outcomes are used to describe performance.

Sensitivity describes the ability of the programme to correctly detect those who have the disease/condition. Poor sensitivity would result in high numbers of missed cases.

Specificity is the ability to identify those who do not have the disease/condition. Poor specificity results in a high rate of over-diagnosis. In screening, it is common to have a series of tests. An initial highly sensitive test, to ensure that all the cases are identified, followed by a second highly specific test to ensure individuals are not receiving unnecessary treatment. Knowledge of the predictive value of a positive result is also highly valuable in evaluating screening, i.e. the likelihood of a positive test result meaning that the individual has the disease. This is often referred to as the positive predictive value (PPV) or power.

2.2.2. Hip surveillance programmes

Internationally hip surveillance programmes have been developed to monitor hip dysplasia in children with CP. The frequency of assessments is normally stratified by GMFCS level. Although assessment intervals and measurement variables differ between different programmes, at minimum they comprise a physical examination to assess passive hip abductor range and hip pain, and a radiological assessment. At this assessment an anterior-posterior radiograph is taken, with the individual in a standardised position. The lateral displacement of the femoral head from the acetabulum in the radiograph is most often estimated using RMP[27].

Given the definition of surveillance programmes, where a population is monitored to detect changes and trends in the dynamics of a condition/disease at a population level, and screening, where the goal is to pre-emptively detect a disease/condition in individuals within a specific population, it is hard to position the hip surveillance programmes into one category. On the one hand they appear like surveillance with routine monitoring, but although learnings and evaluations of the population data are

done, the primary focus appears to be on implementing timely intervention for the asymptomatic individual – which fits into the screening remit.

The world health organisation (WHO) defines criteria for screening tools[96]. Applying these criteria to hip dysplasia in children with cerebral palsy (Table 3), reveals that current programme design would fall short for some criteria.

WHO criteria	*Hip dysplasia in CP*
The condition should be an important health problem	Achieved
There should be a recognisable latent or early symptomatic stage	Achieved
The natural history of the condition, including development from latent to declared disease, should be adequately understood	There is significant heterogeneity in the natural history of the pathology. Hip progression rates vary, whilst risk factors have been identified which allow better prediction of the 'at risk' population they do not completely explain the variability in hip dysplasia progression rates.
There should be an accepted treatment for patients with recognised disease	There is no consensus on what treatments/interventions should be performed at what thresholds. The efficacies of the intervention options are poorly investigated, challenging the ability to define an 'accepted' treatment.
There should be a suitable test or examination that has a high level of accuracy	There is room to improve the accuracy of examination in hip dysplasia. Currently they have relatively high measurement errors associated.
The test should be acceptable to the population	Whilst X-rays and clinical examinations are the standard care and would be deemed acceptable to the population, the protocols for standardising the assessments are not always acceptable to the population. For example, the positioning protocols for X-ray and not always possible to achieve due to deformity or high tone – further reducing the accuracy of measurements derived from these images.
There should be an agreed policy on whom to treat as patients	Largely all children with cerebral palsy are included in hip surveillance programmes, however there is significant variability in frequency of assessment for different individuals.
Facilities for diagnosis and treatment should be available	Achieved
The cost of screening (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole	Hip surveillance programmes are considered cost effective, however the health economics has been under investigated. One economic analysis was found which showed that hip surveillance programmes were likely to be cost effective, measured by quality-adjusted life years (QALY's). Their analysis showed that hip surveillance cost €12,282 per QALY gained. Spain, where the analysis was conducted, have a cost-

	effectiveness threshold of between €20,000-€25,000 per QALY.
Screening should be a continuing process and not a 'once and for all' project	Achieved

Table 3: Table displaying the application of the WHO criteria for screening programmes to hip dysplasia in CP

The efficacy of the clinical tools used to differentiate between those who will develop clinically significant hip dysplasia and those who will not are sub-optimal. However, globally, these programmes have had a positive impact, with fewer patients requiring surgery to salvage fully dislocated hips[24,97]. As with any screening programme, the sensitivity of the screening tool is important. RMP is used as the index for monitoring progression with the threshold at which intervention should take place defined within the programme. Optimisation of monitoring intervals and intervention thresholds are dependent on expected progression rates of hip displacement and the accuracy and precision of the measurement. Measurement uncertainty and heterogeneity of pathology may explain the lack of consensus between surveillance programme protocols.

The original surveillance programmes, as devised in Scandinavia, were focussed on radiographic measurements of the hip. It is worth pointing out that more recent versions of these programmes include other data such as extensive upper and lower limb assessment of range and functional classification. These additional data contribute to inform clinical management of the child including the management of the hip.

Hip surveillance programmes cannot be considered as screening programmes until a test or set of tests with sufficiently sensitivity and specificity are developed. Secondly, there is no clearly agreed pathway for individuals should they receive a 'positive'

diagnosis of hip dysplasia. The lack of consensus may be conflated by the errors in the assessment of RMP, which are so embedded in the majority of studies in the field.

2.3. Validation

Validation is a process of proving that a method, measurement or index quantifies or represents an underlying situation accurately enough for the application. In nearly all circumstances there is no single test that proves, or validates, a technique. Instead, there is a process of evaluation, usually using different approaches to address different aspects of the technique. In the case of validating a clinical test, clinical measurement or index for use in a clinical setting, validation involves both the scientific and clinical communities being satisfied that sufficient evidence has been gathered to apply the index/test clinically with confidence. Defining the point at which the psychometric properties of the application are adequate for the clinical application is challenging. The acceptability of an index may also depend on the clinical culture and environment in which it is applied. For example, there may be some pressure to introduce a solution under non-ideal measurement conditions. X-ray are a familiar imaging modality which orthopaedic surgeons use already and so in spite of the index properties being moderate, and due to the lack of viable alternative techniques, it has become the clinical standard.

2.3.1. Types of validation

There are many different types of validation each supporting different aspects of a method of measurement. Perhaps the clearest cut is criterion validity, where the new measure is compared to something of a known quantity or a gold standard measure. However, in reality, particularly in clinical measurements, it is common that no gold standard exists or it is very hard to define an absolute truth to compare. In these cases we might create a phantom or test object to assess the performance of the measurement where the experimental parameters can be prescribed and the absolute performance of the measurement evaluated.

We often rely on construct or concurrent validity to increase the confidence in the measurement. Construct validity is when the measurement is related to or sensitive to measures that we would expect it to be, and independent of measures that we would expect it to be independent of. Construct validity is often used in situations where physical measurement is not possible, for example validating a questionnaire or scale used to quantify a concept such as pain, or spasticity.

Concurrent validity measures how well a measurement compares to a well-established test or measure. This is often assessed in a levels of agreement study. The strength of the similarity between the measures may be expressed as a simple correlation and/or regression or a Bland Altman plot. The value of concurrent validity in clinical measurement is that it is possible to compare novel measures to clinically accepted, or standard measures which are relevant, understood and trusted. Strong agreement between measurements increases confidence in the novel technique, however it often doesn't describe the absolute performance of the measurement as it

is hard to attribute any disagreement in the two sets of measurements to the psychometric properties of the measurements or error.

Repeatability or test-retest reliability is related to validity. It measures how consistent a measurement is. To test repeatability multiple measurements are required, usually from multiple assessors. A measurement cannot be valid if it is not repeatable, however a repeatable measurement can be invalid if it does not perform in the other domains described. In some cases bias is identified, where the measurements are consistently offset from the 'true' value. If a bias is quantified, it can be accommodated in the clinical application of the technique.

2.3.2. Methods of validation

2.3.2.1 Simulation

Statistical models and mathematical simulation can be used to estimate or predict the outcome of a process without physically implementing the process. The effect or impact of varying different performance parameters of a clinical measurement can also be modelled (a sensitivity analysis). Such models can be useful for understanding or defining acceptable limits of performance for clinical measurement. Methods such as Monte Carlo allow for simulations to be repeated many times to build a probabilistic outcome model. These methods are commonly used in medical physics when defining patient specific radiation dose[98]. The accuracy and validity of these models is dependent on the confidence and accuracy of the input parameters and model decisions. Whilst these can often be modelled in a sophisticated manner, it is challenging to model softer parameters that influence clinical decision making. In this

thesis a Monte Carlo simulation is used to look at the potential impact of measurement uncertainty in hip surveillance programmes at a sample population level.

2.3.2.2. In vitro studies

Phantoms or test objects are used routinely in ultrasound quality assurance, they are representative of certain properties of the clinical situation, i.e. acoustic properties, but are usually not similar in anatomical geometries. They are designed to ensure that simple repeatable checks can be performed to assure the technical performance of the ultrasound. The phantoms are designed so that fixed measurements can be taken, in different modes, to test a variety of parameters on the scanners. Phantoms can also be used to calibrate or tune parameters to optimise images for specific uses[99].

Phantom, or *in vitro*, studies are conducted to assess the technical performance of a clinical measurement, they are often performed as they allow many more sample measurements across a prescribed range than in an in vivo setting. Commonly, *in vitro* studies precede clinical studies, providing initial data and informing the design of the clinical trial. In this thesis the criterion validity of two novel indices of hip displacement are investigated in an *in vitro* set up.

2.3.2.3. In vivo studies

Levels of agreement studies are often conducted *in vivo*, comparing a novel measurement to a trusted clinical standard measurement. Such studies increase the trust in the new measurement and quantify the agreement between the two measurements/techniques. Clinical studies also allow for other parameters to be evaluated, for example the acceptability of the measurement for the patient and

clinicians, the resource or knowledge gap between the research domain and the clinical domain, understanding what equipment needs to be commissioned, or processes adapted or staff that require training/upskilling to conduct the assessments. *In vivo* studies should be designed to capture the range of presentations expected in the target population, for example when conducting a study looking at agreement between a novel measurement of hip dysplasia and the current clinical standard, it is important to ensure that some individuals in the study sample have hip dysplasia and some do not. However, to ensure that there is not bias in the recruitment it is important to ensure that these individuals are not recruited based on their positive or negative status, but recruited regardless of their status. For example, it would be a bias sample if all individuals were recruited while waiting for hip surgery. A large enough sample is therefore needed to ensure that the spread of presentations is captured without bias. There are situations where targeted recruitment is appropriate, particularly where the condition under study is very rare, however generally it is better to blind the study team to as much information about the participants and the trial as possible. In this thesis, to assess the concurrent validity of a novel index of lateral hip displacement (derived from 3D ultrasound), a clinical study was conducted to assess the level of agreement between our novel index and the clinical standard measurement of lateral hip displacement from X-ray.

2.4. Summary

In summary, this chapter outlines several techniques and methodologies that were used in the studies presented. Whilst hip surveillance programmes are titled as such, there are principles of screening programmes incorporated, namely the goal of identifying emerging issues early. The overall collection of data to evaluate the

population and trends in hip dysplasia and treatments, appears to be a secondary goal, but is in line with surveillance principles. This thesis examines several elements of hip surveillance programmes using different techniques. Firstly, the impact on treatment decisions of the design of the radiographic schedule in a typical hip surveillance programme is evaluated by simulation. Secondly, motivated by the measurement errors associated with the clinical measurement RMP, two new indices of hip displacement, derived from 3D ultrasound, are evaluated in an *in vitro* study to assess the criteria validity of the indices. Thirdly a clinical study is conducted to assess the concurrent validity of the novel index of lateral hip displacement, derived from 3D ultrasound. Finally the feasibility and clinical implications of a multi planar quantification of hip dysplasia using 3D ultrasound is investigated.

The collective objective of this work was to investigate the potential of 3D ultrasound in the monitoring of hip dysplasia in children with cerebral palsy.

3. Monte Carlo simulation of a hip surveillance programme

3.1. Overview

In this chapter I discuss a Monte Carlo simulation that was created principally to investigate the impact of error in the measurement of RMP in typical hip surveillance programmes.

The motivation for developing the simulation arose from reflections on the reported poor reliability of RMP measurements and the increasing reliance on a consensus approach to evaluating diverse hip pathology in a heterogeneous population. Whilst standardisation of practice across centres and regions is desirable, the reliability of the underpinning indices may jeopardise the efficacy of hip surveillance programmes using a simple radiographic approach.

A combination of natural history data and reported reliability of measuring RMP from hip X-rays were used to develop a model of a typical hip surveillance programme. The model included different decision thresholds that would trigger a 'recommendation for intervention'. Both the decision thresholds within the simulation and model inputs were varied to look at the impact of varying the frequency and quantity of radiographs/measurements as well as the thresholds at which an intervention would be recommended. The output of the simulation is a probabilistic view of outcome, based on the prescribed inputs.

The core of this chapter was published in the Journal of Orthopaedic Research in 2019, see Appendix 1 for the full article.

3.2. Introduction

Hip surveillance programmes have been adopted internationally to monitor hip development in children with cerebral palsy[24,38,39,97]. Although assessment intervals and measurement variables differ between different programmes, at minimum they comprise a physical examination to assess passive range of movement of the hip abductors and hip pain, and a radiological assessment. At this assessment an anterior-posterior radiograph is taken, with the individual in a standardised position. The lateral displacement of the femoral head from the acetabulum in the anterior-posterior radiograph is most often estimated using RMP[27]. Simply, the index defines the percentage of the ossified portion of the femoral head that is not covered by the acetabulum. The measurement is taken along Hilgenreiner's line. The frequency of assessment is often dictated by the level of function of individuals under surveillance, with individuals with Gross motor function classification levels (GMFCS) IV and V receiving annual or in some cases bi-annual assessments, and individuals who can independently mobilise receiving initial assessment and sometimes no further scheduled assessment. The clinical pathway of an individual is defined by the outcome of each assessment. Thresholds for discharge, continued monitoring and referral for orthopaedic management are defined within each programme, typically a threshold for hip displacement and/or progression of hip displacement, a minimum hip abduction range or the presence of hip pain. There is no consensus on the RMP thresholds, but it is widely accepted that hips with RMP of greater than 33% are either at risk or require intervention, and at 50% migration most clinicians would agree that intervention is required. However, as previously discussed, measurements of RMP are subject to errors in acquisition and analysis. In the acquisition, the content of an anterior-posterior X-ray image depends on both the relative orientation of the subject and the

X-ray source, and the relative position of the femoral and pelvic segments of the hip[75]. In the analysis, variation in the identification of the required landmarks, differentiation of bony borders and tools used to aid the measurement can result in both inter and intra assessor variation which results in a minimal detectable difference (MDD) of approximately 10% RMP[75,76,78].

Considering the large MDDs and the relatively low rates of annual hip displacement progression (7% RMP across GMFCS levels III to V and as low as 1.3% in the GMFCS level III cohort[32]),it is probable that for some hips the *measured* RMP is significantly different from the actual or *true* RMP, i.e. the hip may be mis-classified as 'at risk' when its position is satisfactory, or classified as satisfactory when in fact it is 'at risk'. The size of these groups and the impact of mis-classification are under-investigated.

Since the advent of routine monitoring, total dislocation rates have reduced to almost 0%[24] indicating that, when *true* above-threshold RMP is under-estimated at one radiographic assessment, it is likely that at subsequent assessments an above threshold measurement will be made, i.e. the annual assessment ensures that all significantly displaced hips are eventually detected. However, there is a potential cohort who are falsely indicated for intervention, and who consequently risk undergoing unnecessary treatment. These individuals represent a "hidden" group who would be highly challenging to identify in a clinical study. Hagglund et al[37], eluded to the possibility of this scenario in their study investigating the effect of different RMP thresholds in hip screening. They had a mean reduction in RMP of 10.8%, without operative intervention, in one third of their cohort with an RMP of 33% or more. They state that this should be considered the "minimum value for non-operative improvement, as it is not known whether any of those operated on would also have improved without surgery".

In this chapter a Monte Carlo was simulation created to investigate the influence of uncertainty in the measurement of RMP, specifically, during a prescriptive hip surveillance programme for children with cerebral palsy (GMFCS III – V), is described. We hypothesised that the sensitivity and specificity across the surveillance programme would be high, but that there would be a significant number of cases inappropriately indicated for intervention in a simulated sample population of individuals with cerebral palsy, particularly in those individuals where the underlying rate of progression was low (a poor PPV). The impact of frequency of assessment and number of X-rays per assessment on the diagnostic value of the radiographic schedule were also investigated.

3.3. Methodology

3.3.1. Description of the simulation

The Monte Carlo simulation described in this chapter was developed in Microsoft Excel (Office 365 ProPlus) using Visual Basic for Applications (VBA). It was designed to replicate the radiographic imaging component of a generic hip surveillance programme for non-ambulant individuals with cerebral palsy, and for those who could walk with assistive devices (GMFCS levels III – V) with annual screening between 2 and 8 years of age. Data were all assumed to be normally distributed, the 'NormInv' function in vba was used, with the mean and standard deviation of the distribution varied dependent on what distribution was being modelled, e.g. RMP error, annual progression rates (by GMFCS level), initial presentation data (by GMFCS level). The indication for intervention decision was defined by 3 variable input parameters, a lower intervention threshold, a progression threshold and an upper intervention threshold, IF and OR

functions were used to create a single decision function capturing the 3 variable thresholds. The output of the decision was a Boolean, true or false. The decision function was applied to both the simulated measured and simulated true data points and a classification of 0-3, representing true positive, false positive, true negative and false negative, were assigned to each set of data, representing an individual at each time point. Data 'Types' were defined to capture the sensitivity analysis for a simulated individual within a programme, a simulated cohort and the overall averages from the simulation. This set up allowed for the simulation to be examined at different levels.

3.3.2. Parameter selection

To create representative simulated *true* RMP values (i.e. from simulated measurements that were not subject to error), random data points were generated around a normal distribution defined by the mean and standard deviations of RMP values reported at initial assessment in Terjesen's[32] dataset describing the natural history of hip displacement stratified by GMFCS level (Table 4).

GMFCS level	*Mean initial RMP (s.d.)*	*Mean RMP progression/year (s.d.)*
III	26.5 (10.7)	1.3 (3.1)
IV	26.2 (20.2)	3.9 (4.8)
V	28.6 (24.3)	9.5 (9.4)

Table 4: Natural history hip displacement data by GMFCS level[32]

Simulated cohorts of 1000 individuals per cohort were created for each of these GMFCS levels. For each *true* RMP value in the simulation, a *measured* RMP value was created by adding a simulated normally-distributed measurement error to the *true* RMP value derived from repeatability data published by Craven *et al*[76]. Craven *et al.* published the SEM of a single measurement of RMP as 3.9%, which corresponds to a MDD of 10.8% This value was chosen as a representative error in the simulation

since it was similar to other values of the reliability of RMP in the literature[75,77,78] (Table 5).

Authors	*Reported ICC*	*MDD (%)*
Craven et al[76]	0.93	10.81
Kinch et al[77]		11.00
Cliffe et al[75]	0.96	10.53
Parrot et al[78]*	0.91	11.49

*Table 5: Results from repeatability studies, ICC was quoted in all papers except one. MDD has been computed from data presented in the original articles. Where authors have broken down repeatability data to within and between assessor variations, the between assessor results have been used and their corresponding ICC's quoted. *Data taken from right hip only.*

Progression of hip displacement was simulated according to a normally-distributed random distribution based on the mean and variance of annual hip progression reported by Terjesen[32] (Table 4). In this way, we estimated the *true* and *measured* RMP values in a simulated surveillance programme for children with CP (GMFCS III-IV) between the ages of 2 and 8 years, with annual follow-up.

3.3.3. Simulation decisions

Within the simulation, decision making was based on three thresholds, which collectively defined the 'indicated for intervention' decision. The first, an upper RMP threshold (fixed at 50% RMP throughout), secondly, a lower RMP threshold and finally a progression threshold. Intervention was indicated if the *measured* RMP was greater than the lower threshold and the change in *measured* RMP in successive assessments exceeded the progression threshold, or the *measured* RMP exceeded the upper (50% RMP) threshold irrespective of progression. A sensitivity analysis was conducted to investigate the impact of varying the lower RMP limit and the progression threshold on the decision to intervene. Those children who were indicated for

intervention at any assessment were removed from the simulated programme at that time point.

3.3.4. Simulation stability

To assess the stability of the simulation and establish the optimal number of iterations required, the number of simulation repeats was varied, and the results of the simulation recorded. Total number of radiographic assessments conducted during the simulation was chosen as the summary result. This fluctuates depending on whether a positive or negative decision is made, a positive decision results in no further radiographs, whilst a negative results in continuation to the next time point. Stability in the simulation was defined as the point at which this variable plateaued (variability of less than 2 radiographs) with increasing number of iterations.

3.4. Data Analysis

To test the hypotheses, the indication for intervention decision was assessed using both the simulated true RMP values and the simulated measured RMP values at each time point. Table 6 shows the classification of the results. When the simulated measured and true RMP both satisfied the indications for intervention, the decision was defined as a true positive (the child in the simulation is appropriately indicated for intervention). Similarly, a true negative was defined as an instance where both the measured and true values did not satisfy the indications for intervention (the child is correctly not indicated for intervention). A false positive result occurred when the measured data indicated intervention, but the true data did not (the child is indicated for intervention when intervention should not be indicated). Similarly, a false negative

result was achieved when the measured data did not indicate that intervention was necessary, but the true data suggested intervention was indicated (a child is not indicated for intervention when intervention should be indicated). From these, the sensitivity (Equation 4), specificity (Equation 5), positive predictive value (PPV) (Equation 6), and negative predictive value (NPV) (Equation 7) were computed.

		True data points		
		Positive	**Negative**	**Subtotal**
Measured data points	**Positive**	True positive (TP)	False positive (FP)	Intervention indicated group
	Negative	False negative (FN)	True negative (TN)	No intervention indicated group
	Subtotal	Indicated for intervention group	Not indicated for intervention group	Total number of data points

Table 6: Explanation of the possible categorisation of each of the data points when applying the decision algorithm

$$Sensitivity = \frac{TP}{TP + FN}$$

Equation 4: Equation for calculating sensitivity

$$Specificity = \frac{TN}{TN + FP}$$

Equation 5: Equation for calculating specificity

$$Positive\ predicted\ value(PPV) = \frac{TP}{TP + FP}$$

Equation 6: Equation for calculating the positive predictive value (or power)

$$Negative\ predictive\ value\ (NPV) = \frac{TN}{TN + FN}$$

Equation 7: Equation for calculating the negative predictive value (or power)

To investigate the effect of the lower surgical threshold and progression threshold on the performance of the surveillance programme, simulations were performed at different intervention thresholds for hip displacement and for different rates of hip progression.

3.5. Results

3.5.1. Simulation stability

To assess the stability of the simulation a single output was chosen, the number of simulation cycles were increased until minimal change in the output parameter was observed for the same input parameters. The simulation was taken to be stable when the total number of radiographs across a programme was within 2 radiographs for a whole cohort. This point was reached at 5000 simulation cycles. The cohorts were modelled as 1000 individuals. All further simulations were run 5000 times.

3.5.2. Hypothesis testing

Table 7 shows the sensitivity, specificity, PPV and NPV of the simulated surveillance programmes. Sensitivity is a measure of the proportion of true positive results (i.e. when the simulated ‘true’ data and the simulated ‘measured’ data both indicated a positive result) that are correctly identified. Specificity is the measure of the proportion of true negative results that are correctly identified. Depending on whether the intervention decision included a progression threshold the sensitivity varied from 0.66 to 0.90. Specificity is very high regardless of the parameters of the intervention decision.

	No progression			*10% progression threshold*		
	GMFCS III	**GMFCS IV**	**GMFCS V**	**GMFCS III**	**GMFCS IV**	**GMFCS V**
Sensitivity	0.75	0.82	0.90	0.66	0.80	0.87
Specificity	0.96	0.96	0.96	0.97	0.95	0.95
Positive predictive value (PPV)	0.55	0.78	0.89	0.23	0.63	0.85
Negative predictive value (NPV)	0.96	0.96	0.96	0.97	0.95	0.95

Table 7: Sensitivity, specificity, PPV and NPV across the simulated surveillance programme for each of GMFCS levels III, VI and V. Indication for intervention decision parameters were set at upper intervention limit of 50%, lower limit at 40% and progression threshold at 0% and 10%.

It was hypothesised that there would be a large number of cases that were indicated for intervention as a result of measurement error, and that the proportion of false positives would be greatest in the group with the lowest underlying rate of hip displacement i.e. the GMFCS level III group. Positive predictive power or value (PPV) is a measure of the probability of a positive result being a true positive result, i.e. a PPV of 20% means that 1 in 5 positive results are truly positive. Figure 12 illustrates the influence of the progression threshold and lower RMP limit on the positive predictive power by GMFCS level. Within each GMFCS level, the lower RMP threshold does not have a great influence on the PPV. In the GMFCS III cohort, the PPVs vary between 55% and 70%, depending on the lower RMP threshold, when the progression threshold is set to zero meaning that at this level between 30% and 45%, depending on the lower RMP threshold, of individuals will be incorrectly indicated for surgery according to the simulation. Including a progression threshold in the simulation has a *negative* effect on predictive power particularly in the GMFCS III group. Looking closely at the GMFCS III data, regardless of the lower RMP threshold the PPV drops from over 50% to approximately 20% when a progression threshold of 8% is applied. This implies that under these conditions, only 1 in 5 positive results are likely to be true positive results. The same trend is seen in the GMFCS IV and V data although it is less extreme, this is due to a greater underlying rate of progression in these groups.

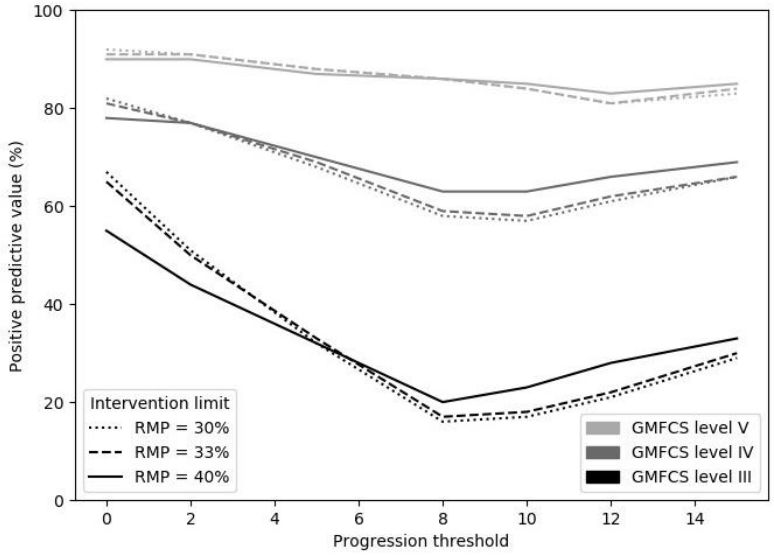


Figure 12: A graph showing the effect of varying progression threshold and lower surgical limit on the predictive power of a positive result.

3.6. Discussion

This simulation was performed to better understand the potential influence of measurement error and decision-making thresholds on the success of a typical hip surveillance programme. The simulation supported our hypothesis that the surveillance programmes would have high specificity, and that due to the relatively large errors in the measurement of RMP compared to typical hip displacement progression rates, a large number of individuals would be indicated incorrectly for intervention by radiographic measurement (RMP). The proportion of individuals falsely indicated for intervention was particularly high in the GMFCS III group where mean hip progression rates was lower than in GMFCS IV and V groups.

3.6.1. Simulation validation

The authenticity of a simulation of a complex clinical process may be open to doubt.. We cannot hope to model the tacit understanding of the clinicians involved nor all the factors influencing a treatment decision. However, the descriptive validity of the simulation can be assessed by comparing the summary simulation results to the published data that underpins the simulation[100]. Table 8 shows the summary results for the simulation (mean and standard deviation of RMP at initial presentation and annual RMP progression), stratified by GMFCS level compared to the published data[32] upon which the simulation is based.

	Initial Presentation				*Progression*			
	Simulation		**Measurements reported in the literature**		**Simulation**		**Measurements reported in the literature**	
GMFCS level	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
III	26.7	10.8	26.5	10.7	1.3	3.1	1.3	3.1
IV	25.6	20.2	26.2	20.2	3.9	4.9	3.9	4.8
V	29.4	24.2	28.6	24.3	9.5	9.3	9.5	9.4

Table 8: Initial presentation and progression summary results (mean and standard deviation) from the simulation stratified by GMFCS level alongside the published data underpinning the simulation

Predictive validity is a measure of how well the results describe data that were not used to inform the simulation. To assess the predictive validity of the simulation the true positive and false positive results were compared to a published analysis of hip displacement rates. Soo *et al*.[33] published proportions of individuals with hip displacement stratified by GMFCS level. Hip displacement was defined as an RMP of greater than 30%. Our simulation defined indication for intervention thresholds in a similar way. Table 9 shows the rates of indication for intervention from the simulation and Soo ***et al's*** hip displacement rates. For each of the GMFCS levels III-V, the simulation produced similar to those published by Soo et al[33]. When no progression

threshold is included, the simulation is slightly less conservative across all GMFCS levels, but when a progression threshold is included in the decision algorithm the simulation becomes more conservative than Soo *et al*'s data.

GMFCS level	**Soo *et al* results**	**Simulation - prog threshold 0%**	**Simulation - prog threshold 10%**
III	43%	45%	26%
IV	69%	71%	62%
V	89%	93%	92%

Table 9: Table showing Soo et al. hip dysplasia rates and simulation indication for intervention rates by GMFCS level. The upper RMP threshold was set at 50% throughout, lower RMP threshold set at 40% and the progression threshold set at 0% and 10% RMP.

3.6.2. Clinical implications

Although, under certain conditions, the sensitivity of hip surveillance programmes appears to be moderate, the design of surveillance programmes means that children with hip displacement who are missed at a single assessment, will most likely be detected at the next assessment without clinically-significant amounts of progression in the interval, thus increasing the detection rate of hip displacement. Due to the large number of true negatives in a surveillance programme, the use of sensitivity and specificity alone as a measure of programme performance may be flattering. When a positive result does occur, it is important to consider how likely it is that this result is a true positive – this is described by the PPV. High PPV is important when interventions with potential morbid outcomes are being considered.

To better understand the impact of these results they are framed in the context of the UK population. The incidence of CP in the UK is approximately 2.2 for every 1000 live births[6]. Assuming 750000 live births in the UK each year this equates to 1650 children with Cerebral Palsy every year. If a surveillance programme similar to the one modelled here were adopted nationwide we would expect 1650 children to be

introduced to the programme each year. Reid *et al*[101] published data on the distribution of GMFCS levels within the Victorian cerebral palsy register. Using these distributions the number of children at each GMFCS classification born each year in the UK can be estimated (Table 10).

GMFCS	Distribution	Number of individuals born each year in the UK
I	34.2	564
II	25.6	422
III	11.5	190
IV	13.7	226
V	15.6	257

Table 10: Estimated distribution of children with cerebral palsy by GMFCS level born in the UK each year

Based on the results of the simulation, it is estimated that in the region of 45 individuals with GMFCS level III would be falsely identified for intervention due to variability in the measurement of RMP. This number is based on taking a single X-ray, if a progression from a previous X-ray were to be included this number increases to a worst case scenario of approximately 70 due to the potential for error to be introduced in the calculation of RMP from two X-rays (Figure 12).

To investigate the impact of a false positive result, a further time point was simulated to quantify the number of individuals, falsely indicated for intervention whose hip migration would progress enough in the course of the following year to pass the threshold for indication for intervention. In this way, it is possible to differentiate those who were simply, prematurely indicated for intervention and those who were falsely indicated for intervention and would still not have been indicated for intervention at the subsequent time point. Figure 13 shows the premature indication for intervention rate within the false positive group by GMFCS level. In the GMFCS V cohort we can see that 60%-80% of those falsely classified as indicated for intervention were merely pre-

emptive. However, in the GMFCS III group only 10%-35% of the total false positive group were pre-emptive, indicating that the majority of those falsely indicated for intervention were not indicated for intervention at the next time point.

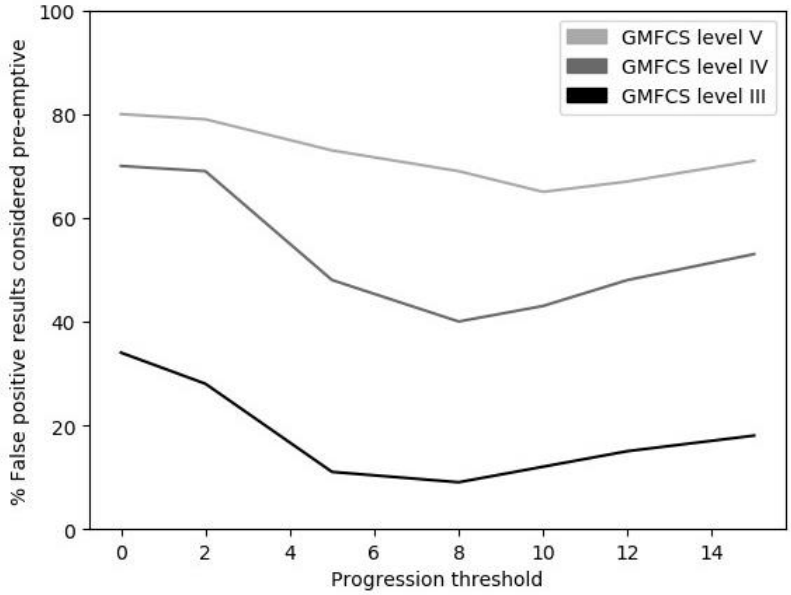


Figure 13: A graph showing how 'pre-emptive' indication for intervention rates vary with GMFCS level and applied progression threshold.

To further investigate the total impact of misclassification, a survival analysis was conducted to illustrate the cumulative chance of a false positive (solid lines) or false negative (dotted lines) classification by GMFCS level across the duration of the simulated surveillance programme Figure 14. The cumulative correct classifications are shown as hashed lines. The simulation parameters were set at RMP lower limit = 40% and progression threshold = 10%. The chance of misclassification at any time point is mutually exclusive and therefore the total chance of misclassification for an individual participating in a complete surveillance programme is the summation of the chances at each time point.

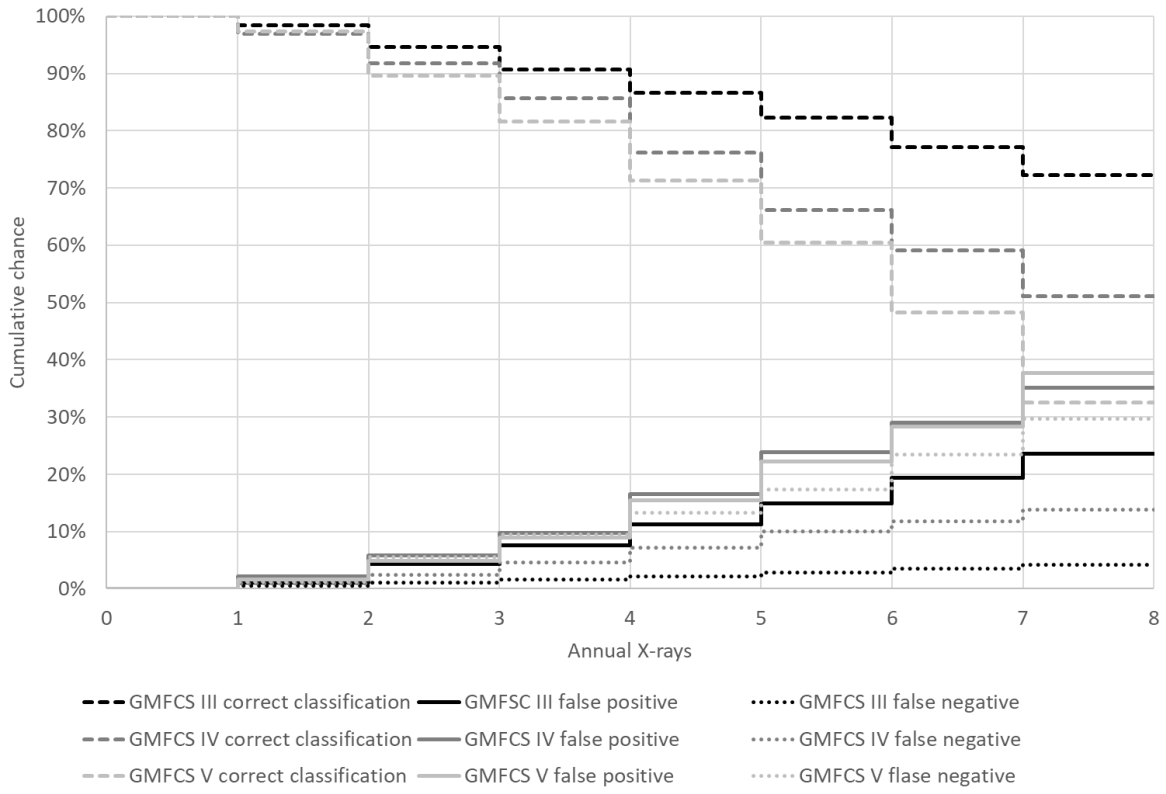


Figure 14: Survival analysis illustrating the cumulative chance of misclassification in a hip surveillance programme by GMFCS level

Despite having poor positive predictive power, the GMFCS level III cohort have the greatest cumulative chance of a correct classification. This is because a large proportion of the cohort were correctly classified as true negatives. At each increasing time point the chance of a misclassification increases. The chance of misclassification at each X-ray is independent of the previous X-rays, but not constant. At each increasing time point the true RMP is likely to progress closer to a decision threshold, where the chance of measurement error resulting in a misclassification is higher.

Despite GMFCS level V having a reasonable PPV (Figure 12), the cumulative chance of a false positive result is high (Figure 14). Although these results initially appear contradictory, they are explained by the high positive rate for the GMFCS V population at some point in the programme.

This survival analysis highlights the overall dynamics of the impact of measurement error, but should not be over interpreted as a standalone analysis, just as the overall sensitivity, specificity PPV and NPV can be misleading. The combination of these summary metrics, alongside cumulative analysis are important.

Where the rates of hip displacement are slower, applying a progression threshold increases the chance of misclassification (Figure 12). Measuring progression requires comparing radiographs from different time points, typically one year apart. Therefore, there are two instances where measurement uncertainties are introduced. In the GMFCS III cohort, when a positive result is indicated it should be treated with caution, particularly if progression is considered in the clinical decision-making.

The American Academy of Cerebral Palsy and Developmental Medicine (AACPDM) guidelines[39] suggest bi-annual follow up for most of this cohort until the age of 5 and to continue with bi-annual screening unless stability has been observed for 2 years (stability is defined as an RMP<30% or less than 10% change in RMP across 2 years). However, increasing the frequency of assessment will increase the chance of misclassification. The AACPDM ‘worst case scenario’ i.e. biannual radiographic screening was simulated for GMFCS levels IV and V, the schedule was simulated from 2 years to 8 years to allow for comparison between the annual screening programme. For comparison to other analyses the intervention limit was set to 40% and tested with progression thresholds of 0% and 10% (Figure 15). The AACPDM guidelines states RMP >30% as the orthopaedic referral criteria, the impact of this lower intervention limit is displayed as the ‘x’ markers in Figure 15.

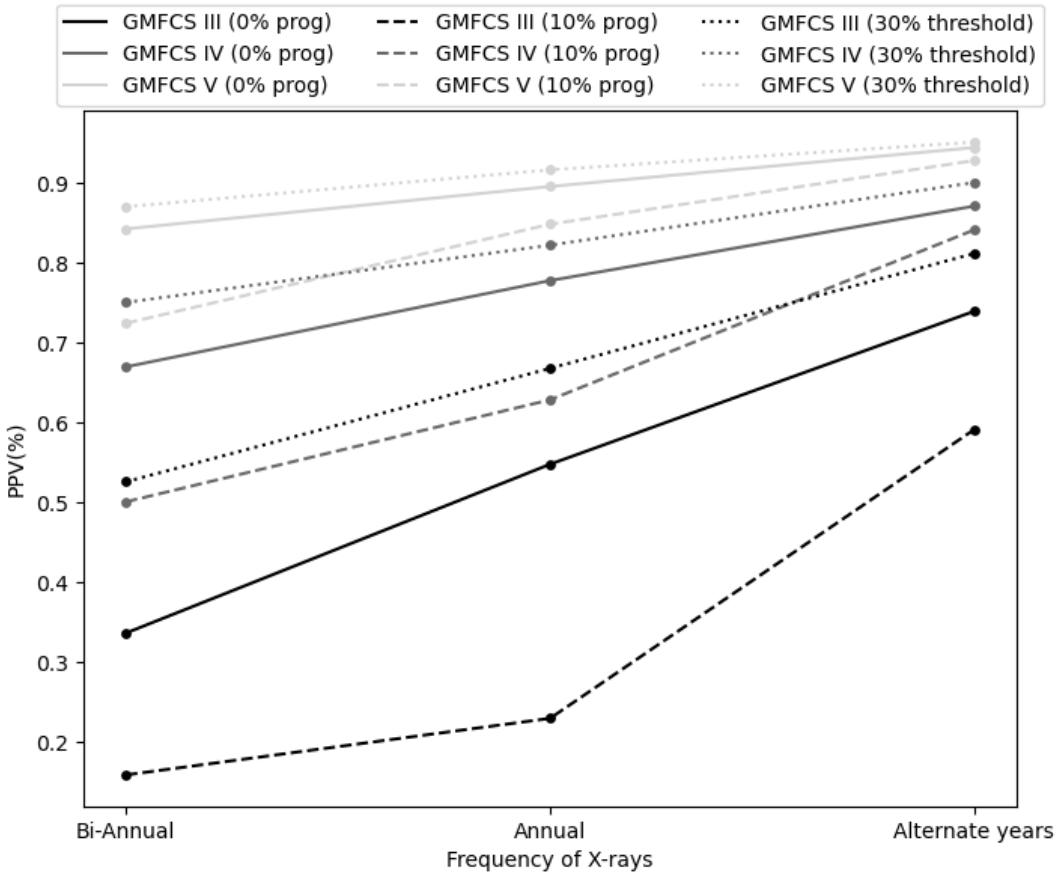


Figure 15: Investigating the impact of assessment frequency of positive predictive power of assessment

On average the progression of hip displacement in individuals with cerebral palsy is below 10% RMP per year, which is comparable in magnitude to the measurement's MDD. Sampling more frequently will minimise the amount of true progression between monitoring points, increasing the chance of measurement error influencing the decision. Conversely if the frequency of the sampling is reduced, the impact of measurement error is reduced, particularly in the GMFCS III cohort (Figure 15).

It is important to ensure that monitoring intervals are optimised with regard to the expected progression rates of individuals to limit misclassification rates, whilst ensuring that individuals with high progression rates are detected in a timely manner.

The impact of measurement error could also be reduced if the average of multiple measurements were taken at each time point - a principle known as regression to the mean. This was simulated in a similar way (Figure 16). The lower indication for intervention threshold was set at 40% throughout and the progression threshold set at either 0% or 10%.

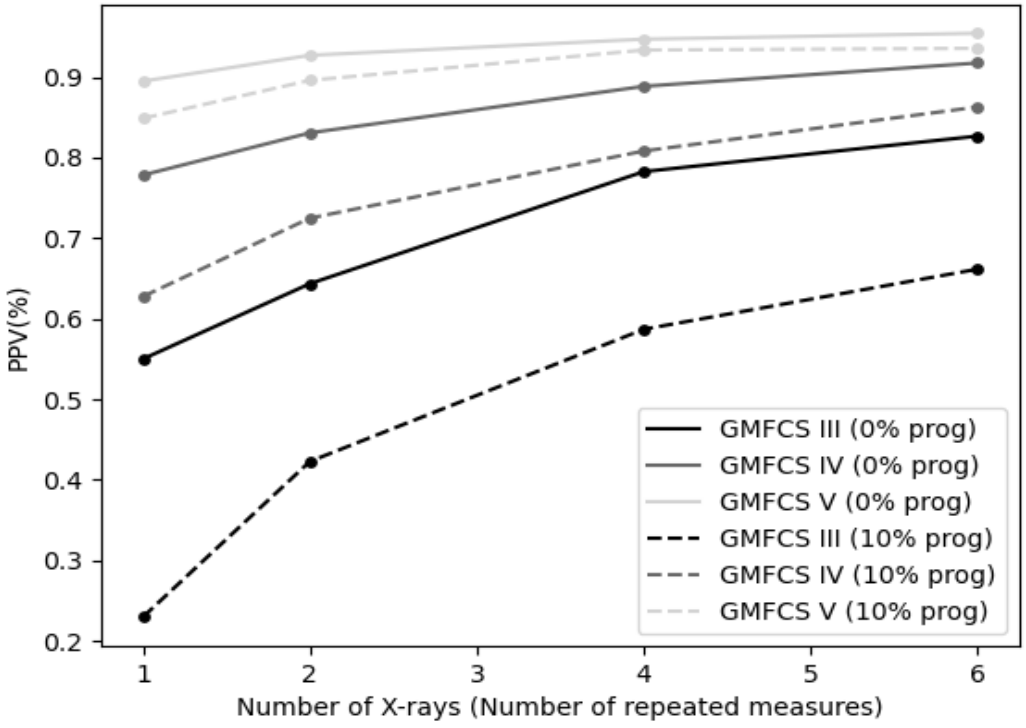


Figure 16: Investigating the impact of multiple measurements at each annual assessment

To optimise the efficacy of hip surveillance for the GMFCS III population, the frequency of assessment could be reduced to alternate years and two x-rays taken at the assessment, with RMP measurements averaged. This would be particularly beneficial if a comparison to the previous X-ray is desired. The positive predictive value of the RMP measurements also increases for both the GMFCS level IV and V populations with repeated measurements at an increased time period between X-rays however in this population there is a greater chance of significant progression with a two-year gap between assessments. Alternatively, the use of a non-ionising alternative imaging technique, such as ultrasound, would enable the benefits of repeated measures, greater confidence in the accuracy of the measurements, without the requirement to compromise assessment frequency.

Whilst hip surveillance programmes are not screening programmes, some of the principles of screening programmes can be applied. In both screening and surveillance, it is important to ensure that no individual who truly has the condition is missed. A sensitive, yet economic and simple to administer test is adopted. In a screening programme a secondary highly-specific test is then applied to confirm a positive result. Once there is a positive radiographic result, particularly in the GMFCS level III cohort, it may be advisable to seek further imaging which better captures the morphology of the acetabulum and femoral head.

3.6.3. Limitations

This is a mathematical model of a clinical scenario, and therefore has limitations. Perhaps the most significant is the assumption that the underpinning data is normally distributed (see for example Terjesen[32]). Secondly, the value for MDD was taken as an average of results of reliability studies from the literature (Table 2). In practice, MDD

may vary according to the experience of the local team and their protocols for position of the patient for radiography. Finally, Hermanson *et al*[29] showed that age and increased RMP at initial presentation are risk factors for progression of hip displacement. This was not included in our simulation.

Error data were assumed to be normally distributed with zero systematic bias, however we know that a systematic error due to X-ray absorption in different tissues is likely to exist. Bone absorbs X-rays much more than the surrounding tissues, resulting in high contrast images of the skeleton in the adult. However, in the infant, the bones of the hip are largely cartilaginous, with the development of ossified bone occurring as the child matures[22]. Systematic measurement errors in planar radiographic imaging are likely as the contrast between non-ossified bone and other tissues is less clear. Unfortunately, we do not know enough about the development of ossification in the hips of children with CP to quantify this error and its potential influence on RMP measurements. There is a potential variation in reliability of RMP measurements with age[78], however Craven *et al* did not find a significant difference in the repeatability of RMP between their age bands[76]. An age-dependent error function was not included in our model due to insufficient published data. The SEM of measurement used in this simulation is in line with reported values in the literature which are based on measurements taken across a full age spectrum of children.

3.7. Conclusion

These simulations indicate that individuals may get indicated for intervention that don't need it due to measurement error in planar radiography. The size of this group is

influenced by the underlying rate of hip displacement and parameters used to define an intervention decision (critical levels of hip displacement and progression). There is an increased risk of misclassification when measurement from radiographs at successive time points are compared, particularly when the underlying (“true”) rate of hip displacement is low. This is because errors taken at two different time points add. However, both reducing the frequency of assessment and averaging multiple measurements at each time point reduce misclassification, with the greatest reductions seen in the GMFCS III cohort where it is likely that the benefits of reducing the assessment frequency to alternate years out weight the risk of any abnormally rapid progression in this population.

In annual screening indications for intervention from planar radiographs, in individuals categorised as GMFCS III, should be treated with caution and further investigations should be considered.

4. Development and performance of indices of hip displacement derived from 3D ultrasound - an *in vitro* study

4.1. Overview

In this chapter an investigation of the use of 3D ultrasound for the assessment of hip migration using an *in vitro* system is reported. The aim of this work was to define and evaluate indices of displacement in both the sagittal and coronal planes. Hip phantoms were 3D printed from computed tomography (CT) scans of patients' hips. The segments of the hip phantoms could be manipulated so that the femoral and pelvic segments could be presented in different relative positions and orientations.

For the purpose of the study, two indices of hip migration were developed. The first index was defined as lateral head coverage (LHC) and describes the proportion of the femoral head that is covered by the acetabulum in the coronal plane. The image is acquired with the hip in a 'side lying' position. The second index describes the position of the femoral head relative to the anterior border of the acetabulum in the sagittal plane (femoral head posterior position ratio (FHPPR)). These images were acquired with the hip in a 'supine' position. Both indices were then tested to evaluate their sensitivity to in-plane and out-of-plane displacements as well as anatomical rotations about each of the flexion/extension, ab/adduction and internal/external rotation axes. The controlled setting was designed to evaluate the potential of 3D ultrasound, in the absence of uncertainties associated with clinical investigations.

The *in vitro* set up involved designing and manufacturing a mechanical rig with 6 degrees of freedom, three rotational and three translational. The pelvic segment was

secured to a base plate and the femoral segment secured into the rig and translated or rotated relative to the pelvic segment. This experimental set-up was then submerged in a water tank and the hips were manipulated in a methodical controlled manner about each of the degrees of freedom in turn. 3D ultrasound images of the phantoms in different controlled positions were acquired. This allowed for the sensitivity to each of lateral displacement, posterior displacement, flexion/extension, ab/adduction and internal/external rotation, in both supine and side lying positions, to be evaluated, for both LHC and FHPPR. As hypothesised LHC and FHPPR were highly sensitive to displacements in their plane of measurement, however LHC was also sensitive to sagittal plane, or posterior, displacements. LHC was largely insensitive to anatomical rotations with the exception of ab/adduction where for every 1º of adduction LHC reduced by 0.5%. FHPPR was largely insensitive to rotations and out of plane translations.

The repeatability of the image analysis, both within and between assessor variance, was investigated for the LHC index. The intra and inter operator variation was very similar to those reported for equivalent, widely used, index measured from planar X-ray (RMP).

Finally this chapter concludes with a discussion about the limitations of the techniques used and the clinical implications of the findings. Where possible these findings were subsequently tested in an *in vivo* setting (chapters 5 and 6).

4.1.1. Acknowledgements

Liam Johnston (LJ) contributed significantly to the design and development of the hip phantoms (section 2) and Michael Jeffryes (MJ) to the mechanical rig design (section

5). They both contributed to the acquisition of the images and were assessors in the repeatability study. The hip phantoms were printed by medical physics (GSTT) and the mechanical rig was manufactured by the team at the GSTT mechanical workshop.

4.2. Introduction

In this chapter an investigation of the use of 3D ultrasound for evaluation of hip migration in an *in vitro* setting is reported. 3D ultrasound is a technique use in foetal medicine and abdominal medicine for the identification of abnormal soft tissue features. Recently, freehand ultrasound has been used to evaluate bony morphology (hip position, femoral anteversion)[87,102].

Ideally, a new measurement methodology should be validated against an available gold standard. In the case of the measurement of hip migration, CT imaging could serve as a validating measure. However, children with CP, in general, do not have the morphology of their hips investigated by CT scan (which would afford a direct comprehensive comparison of indices of hip migration with 3D ultrasound). Even if CT scanning was used clinically to evaluate the hip in CP, it would be impractical and unethical to conduct certain comparisons of the methodologies (say, for example the sensitivity of the results to limb position). *In vitro* studies evaluating a new technique are useful when a gold standard to which to compare is not available clinically. Further, *in vitro* studies permit a systematic investigation of sensitivity to error that would be difficult to do *in vivo*.

Hips in CP are assessed routinely by planar X-ray imaging. The most common method of quantifying lateral hip displacement is by computation of a ratio of lengths made from the radiograph. The Reimer's migration index (RMP) measures the proportion of the femoral head that protrudes past the lateral border of the acetabulum, at a fixed level. 3D ultrasound is unable to view some of the landmarks used in the computation of RMP, so instead a complementary index was developed based on the coverage of

the femoral head from the acetabulum. Assessing the results of a new method and those of the current clinical standard is known as concurrent validity[103].

To our knowledge, an *in vitro* simulation of the hip complex for evaluation by 3D ultrasound has not previously been developed. However, the construction of a test rig for the purpose of measurement validation is common[86,104–107]. Ultrasound phantoms are commonly used to assess the accuracy of measurements taken from ultrasound, and often designed for specific use cases[86,107]. Ultrasound phantoms are used in two ways. The first is a test object that is not anatomically representative, these phantoms are used to assess the accuracy of the ultrasound system. These are often used when calibrating the machine or running quality assurance tests on the ultrasound system. The second type of phantom is one that gives a faithful representation of the anatomy, where features are incorporated to best mimic the clinical situation. These phantoms are used to test the performance of derived clinical measurements.

The potential for clinical utility of 3D ultrasound in monitoring of hip development in children with CP will, in part, be dependent on sensitivity of the chosen indices to anatomical positioning and true displacements (criterion validity). An optimal index would be insensitive to the relative anatomical positioning of the hip but highly sensitive to genuine displacements of the femoral segment relative to the pelvic segment of the hip. Acceptable levels of intra and inter-operator variance in the analysis of the images is also critical for the efficacy of a clinical imaging technique (test-retest validity).

The following requirements were developed for the *in vitro* set-up.

1. Anatomically realistic models of the hips.

2. Anatomical rotations of the in vitro system are representative of the population under investigation.
3. Posterior and lateral displacements could be simulated.

4.2.1. Hypotheses

It was hypothesised that both indices of displacement would be insensitive to variation in relative orientation of the femoral and pelvic hip segments, within +/- 20 degrees of 'neutral' in each anatomical axis (internal/external rotation, ab/adduction, flexion/extension).

It was hypothesised that an index developed to measure lateral displacement would be directly proportional to medial-lateral translation of the femoral head relative to the pelvic segment but would be less sensitive to posterior translations. Similarly, it was expected that an index developed to measure posterior displacement would be directly proportional to posterior displacement of the femoral segment relative to the pelvic segment but insensitive to lateral translations.

It was hypothesised that the indices developed would have similar repeatability to the clinical standard measure of hip migration from planar X-ray, RMP.

4.3. Development of manufactured hip models

The hip phantoms were created from a CT scan of a boy with CP aged 13 years old. The scans were acquired for clinical purposes. Written consent to use this scan for research purposes was gained from the child's father. The scan was acquired at 100

kV and 135 mA with an axial slice thickness of 1 mm. Slicer 4.10.1 was used to segment the image volume. Segmentation was done using a combination of automatic thresholding based on the Hounsfield units (a measure of radiodensity) of each voxel and manual segmentation. The femoral and pelvic sections were segmented separately. The segmentations were them smoothed and exported as .STL files to Meshmixer 3.5 (Autodesk Inc.). The meshes were simplified to reduce the file size to facilitate import to SolidWorks Student Edition 26.3.0.63 (Dassault Systèmes) for modification. Visual inspection of the surface details was used to ensure that they were not compromised by the reduction in mesh points.

SolidWorks Student Edition 26.3.0.63 was used to modify the hip models to facilitate mounting of the phantoms into the mechanical rig. A stand that allowed for freestanding of the pelvic segments in both supine and side-lying positions was added to the pelvic segments. The femoral segments were altered to include a mounting block positioned to ensure that the femoral segment would pivot about the centroid of the femoral head when mounted in the mechanical rig. Guides were designed to allow for the original relative position of the femoral and pelvic segments to be returned to. These could be removed to allow for the segments to be manipulated during experimentation but were used to ensure there was no drift or unexplained movement of the rig throughout testing. The completed models were exported as .STL files and 3D printed using a Polyjet Objet 500 Connex1 (Stratasys Ltd) printer and Polyjet VeroWhitePlus RGD835 material (Figure 17).

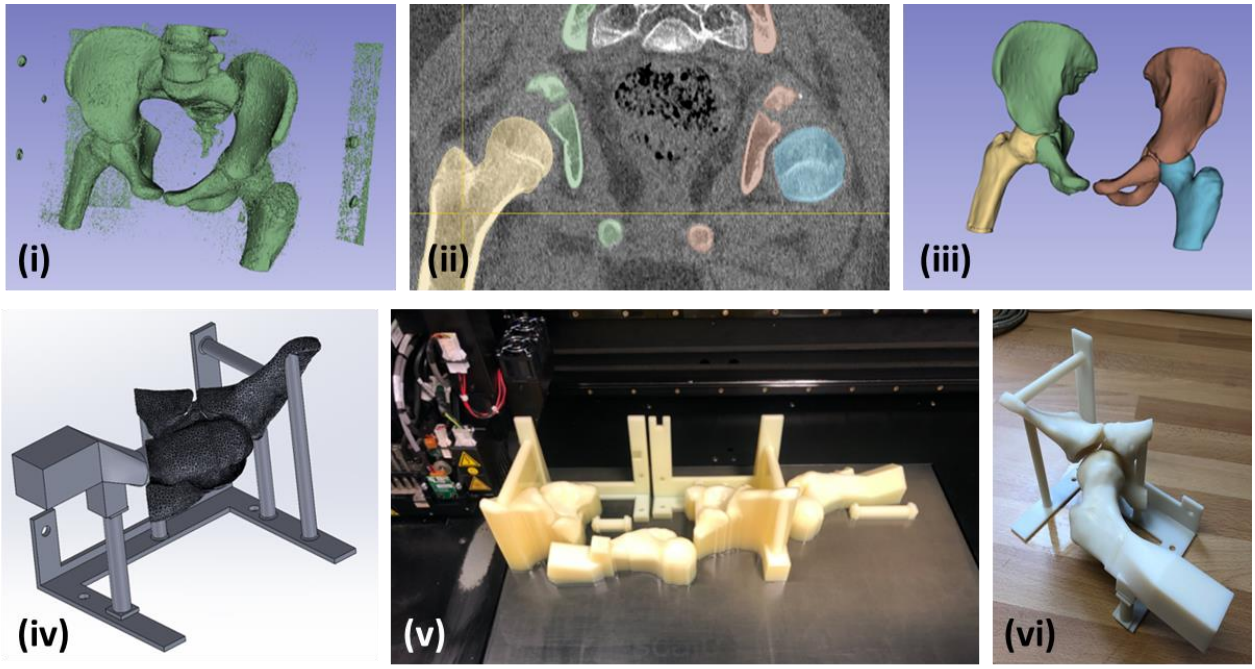


Figure 17: (i) automatic segmentation using thresholding; (ii) plus manual segmentation, split and refined; (iii) smoothed meshes for each segment; (iv) a complete model ready to print; (v) the printed phantoms on 3D print bed; (vi) finished right hip.

4.4. Materials and Methods

4.4.1. Design and development

To assess how the indices performed across a wide range of hip position it was necessary to design a rig that would move the segments by known distances/rotations. The rig needed to be manipulated whilst the phantoms were submerged in a water bath to allow for ultrasound images to be acquired.

A 6 degree of freedom mechanical rig, comprising a gimbal system suspended below an external frame supported by 4 tripods, was designed to allow rotational and translational movement of the femoral segments of the phantoms relative to the pelvic segments. The gimbal system had a mount fixed to one of the arms to hold the femoral segments, ensuring rotations in each plane about the centroid of the femoral head

Figure 18 (i)). Each of the 3 gimbal arms was manufactured (Mechanical workshop at GSTT) to have fixed positions at 5-degree intervals or could be tightened to be fixed at any position.

The gimbal rig was suspended from a frame constructed from steel bar, allowing translation of the rig in two horizontal planes (Figure 18). Corner mounts were designed to attach to tripods, allowing translation in the perpendicular plane.

There were four objectives of this study:

1. Define indices of hip displacement in both the coronal and sagittal planes.
2. Assess the sensitivity of these indices to relative anatomical rotation of the femoral segment of the hip relative to the pelvic segment.
3. Assess the sensitivity of the indices to displacement of the femoral segment relative to the pelvic segment of the hip both in the plane of measurement and in the orthogonal plane.
4. Assess the reliability of the indices.

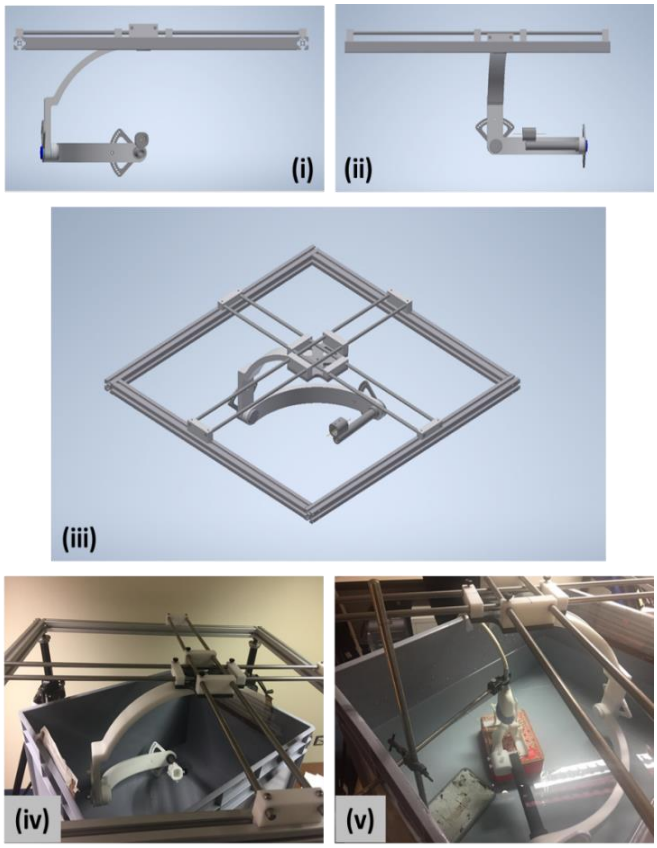


Figure 18: (i) CAD render of mechanical rig (projection view), (ii) Same as image (i) projection view 90 degree rotated, (iii) Isometric view of CAD render, (iv) Photograph of gimbal set up hanging from translational frame. (v) Photograph of experimental set-up with a hip phantom fixed and ultrasound probe positioned.

4.4.2. Development of hip migration indices

In children with cerebral palsy, the hip migrates predominantly in the lateral direction but may have a component of posterior or anterior displacement. In the clinical situation, absolute measurements of displacement are not taken. Instead, the ratios of the lengths of features within the images are recorded. For example, the RMP is a ratio of the distance the femoral head protrudes past the lateral border of the acetabulum, to the diameter of the femoral head. Both measurements are taken in the coronal plane along Hilgenreiner's line (Figure 6). It is presumed that the method of ratios accommodates individuals of different sizes and is less vulnerable to errors in scaling of images.

4.4.2.1. Coronal plane index - Lateral head coverage (LHC)

Lateral head coverage (LHC) was developed to describe the coverage of the femoral head by the acetabulum in the lateral plane. Typically, there is a significant element of lateral displacement of the femoral head relative to the acetabulum in hip dysplasia[30]. Reimer migration index (RMP)[27], from anterior-posterior radiographs, is the current clinical standard measurement used for assessment of hip dysplasia[39,108,109]. LHC is a similar, but not completely comparable, measurement to RMP. It is not possible to directly measure RMP from ultrasound volumes due to the construction of the ultrasound images. RMP relies on the identification of Hilgenreiner's line, to provide the level at which the measurements are taken. In order to identify Hilgenreiner's line, the inferior aspects of the triadiate cartilages are required. In ultrasound imaging these points are not identifiable as they are positioned in the shadow cast by the femoral heads.

LHC was derived from 3D ultrasound images taken with the probe on the lateral aspect of the hip. When acquiring the images in the *in vitro* setting, the hip phantoms were positioned into the "side-lying" position. Considering the clinical setting and ensuring ease for standardising protocols, the long axis of the probe was aligned with the long axis of the pelvis and the notch on the probe (which indicates scan orientation) was orientated superiorly. When acquiring the images *in vivo* the greater trochanter was identified, and the probe moved superior-posteriorly centring over the femoral head so that a suitable view of the femoral head could be found.

The image volume was captured and the perpendicular slices correlating to the maximal cross-sectional of the femoral head area in both the coronal and sagittal planes were selected for analysis. A best-fit sphere was manually fitted to the lateral curvature in both planes and the femoral head diameter (FHD) estimated *d* (Figure

19). The coronal plane slice is used to take the measurements. The superior-lateral border of the acetabulum was identified and the distance in the medio-lateral direction between this acetabular border and the lateral border of femoral head measured *e*. The ratio of the two measurements (*d* and *e*) was taken and deducted from 1 to give an estimate of the proportion of the femoral head that is covered by acetabulum, referred to as lateral head coverage (LHC), Equation 8.

$$LHC = 1 - \left(\frac{e}{d}\right)$$

Equation 8: Lateral head coverage (LHC) an index for quantifying the femoral head coverage in coronal plane. 'd' is the estimate diameter of the femoral head and 'e' is the distance in the medio-lateral direction between this acetabular border and the lateral border of femoral head.

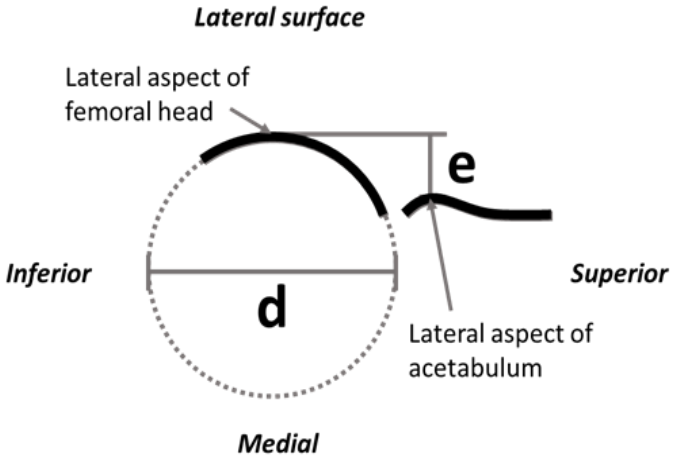


Figure 19: Schematic of hip morphology and bony landmarks used to measure lateral head coverage

4.4.2.2. Sagittal plane index - femoral head posterior position ratio (FHPPR)

FHPPR was derived from the 3D ultrasound images taken with the probe placed on the anterior aspect of the hip. In the *in vitro* set up, the images were acquired with the phantoms in the "supine" position. Again, the long axis of the probe was aligned with the superior-inferior axis of the pelvis and the notch on the ultrasound probe was

positioned to point superiorly up the superior-inferior axis of the pelvis. In the *in vivo* setting, the images were acquired with the subject in supine lying with the probe orientated to be parallel to the superior-inferior axis of the pelvis. The ultrasound volumes were analysed using a similar method to that used for calculating LHC. The slice, in both sagittal and coronal planes, with greatest cross-sectional area of the femoral head was chosen for analysis. A 'best fit' sphere was fitted to the anterior curvature of the femoral head in both the sagittal and coronal planes. The diameter of the sphere, FHD, was measured *d*, within the same slice, the posterior-inferior border of the anterior aspect of the acetabulum was identified and the distance in the anterior-posterior plane to the centre line of the femoral head was measured *a*, Figure 20. Like LHC, FHPPR was constructed as a ratio with femoral head diameter providing a 'normalising' measurement, Equation 9.

$$FHPPR = \frac{a - 0.5d}{d}$$

Equation 9: Femoral head posterior position ratio (FHPPR) an index for quantifying the femoral head positioning in the sagittal plane relative to the anterior border of the acetabulum. The index is reported as a ratio, normalised to femoral head diameter. 'd' is the diameter of the sphere and 'a' is the distance between the posterior-inferior border of the anterior aspect of the acetabulum the centre line

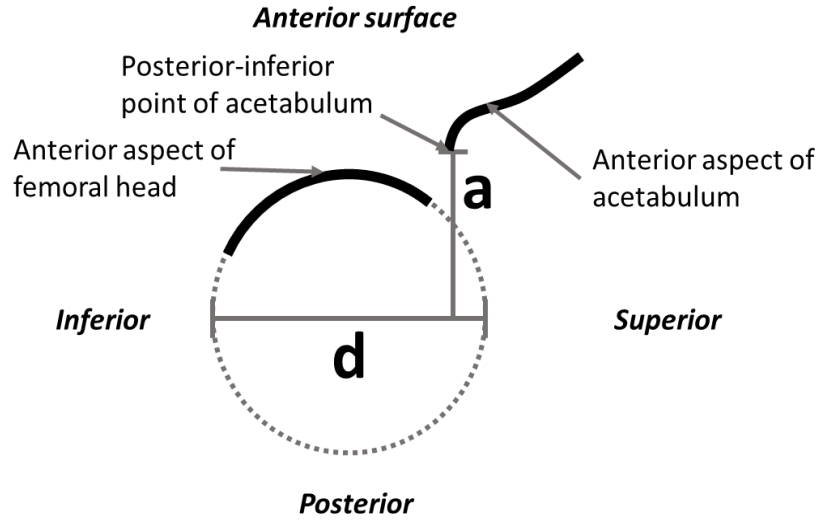


Figure 20: Schematic of hip morphology and bony landmarks used to measure femoral head posterior position ratio (FHPPR)

4.4.3. Methods

The mechanical test rig was used to manipulate the femoral segment of the phantom hips relative to the corresponding pelvic segment phantom. The rig was suspended over a large water bath (800x600x420 mm^3), the pelvic segment was fixed in position in the water bath, mounted on a base plate whilst the femoral segment was mounted into the rig. The guide stands were used to ensure a consistent starting position, which replicated the relative segment positions from the CT scan.

A Voluson scanner with a GE Healthcare RM6C 3D sector scanning probe was used for the image acquisition. The ultrasound probe position was fixed relative to the pelvic segment. The roll axis of the probe was parallel with the long axis (superior-inferior) of the pelvis. The pitch axis of the probe was maintained parallel to the water surface whilst the yaw axis was fixed perpendicular to the water surface (Figure 21).

Four experimental set ups (defined below) were tested:

i. Left hip phantom – side lying
ii. Left hip phantom – supine
iii. Right hip phantom – side lying
iv. Right hip phantom – supine

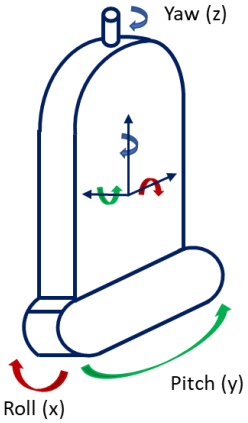


Figure 21: A schematic of an ultrasound probe with the axis (pitch, roll and yaw) highlighted.

4.4.3.1. Side lying set up (i & iii)

Experiments were conducted to test the sensitivity of LHC to different hip positions and orientations. The pelvic segments were secured to the base plate in the side lying position. The femoral segments were secured into the rig, using the mounting blocks that had been incorporated into the 3D print. The mounting blocks were designed for each of the phantoms to ensure that the femoral segment rotates around the centroid of the femoral head. The hips were positioned into the printed positions using the printed guides and the rig set to neutral in all rotations and in the middle for each of the translational degrees of freedom.

To test the sensitivity of LHC to the relative orientation of the femoral and pelvic segments the rig was manipulated to apply prescribed degrees of flexion/extension, ab/adduction and internal/external rotation at 5-degree increments. Images were acquired between 10° extension and 45° flexion, 40° abduction to 40° adduction and 40° internal rotation to 40° external rotation All image volumes were stored for later analysis. Between experiments assessing the sensitivity of the indices to rotation in each axis, the hips were returned to the starting positions and imaged.

The CT scan showed the left hip was internally-rotated. This level of rotation limited the range of motion that was possible without the femoral segment being mechanically restricted by the pelvic segment. To ensure that the experiment was conducted over the full range of positions an approximate anatomical neutral position was estimated by re-slicing the CT scan and calculating the correction angle required to place the femoral segment into a neutral position relative to the pelvic segment. An external rotation of 40º was applied to the left hip throughout testing. This 'corrected' position was defined as the starting position for this hip phantom. The correction allowed the sensitivity of the hip indices to be calculated over a wide range of hip rotation.

Once the anatomical rotations were completed, the sensitivity to translational movements, was assessed. The femoral segment was displaced in the lateral direction by a total of 20 mm in 1 mm increments. It was then displacement medially by 5 mm, in increments of 1 mm. Once Image volumes were collected for each translation in the medio-lateral plane the hip phantom was returned to the starting position. The femoral segment was then translated, again in 1 mm increments, in the anterior-posterior plane by sliding the gimbal system along the frame. In order to posteriorly displace the right hip 5 mm of lateral displacement was first applied as without this compensation the femoral head would be mechanically restricted by the posterior wall of the acetabulum almost immediately.

4.4.3.2. Supine set up (ii & iv)

These experiments were conducted to test the sensitivity of FHPPR to different hip positions and orientations. Similar to the previous set-up the pelvic segments were secured to the base plate, this time in the "supine" position. The femoral segments

were secured into the rig, using the mounting blocks. The hips were positioned into the starting positions using the printed guides and the rig set to neutral in all rotations and in the middle for each of the translational degrees of freedom.

The sensitivity for FHPPR to the relative orientations of the femoral and pelvic segments was tested first, following the same methodology described for the “side-lying” experiments. The femoral segment was manipulated about each of the 3 axes of rotation in 5-degree increments. Image volumes were acquired. All ultrasound volumes were stored for later analysis. Between performing the sensitivity experiment in each axis of rotation, each axis of rotation, the hips were returned to the starting positions and imaged.

Once the anatomical rotations were completed, the sensitivity to translational movements, was assessed. In increments of 1 mm, the femoral segment was translated relative to the pelvic segment, by adjusting the tripod stands to move the entire rig (with the femoral segment attached) whilst the pelvic segment remained in position attached to the base plate. The femoral segment was displaced in the posterior direction by a total of 15 mm in 1 mm increments. It was then returned to the starting position and then displace medially by sliding the gimbal system along the frame medially by 5 mm, in increments of 1 mm. To allow for these ranges to be tested, it was necessary to initially displace the right hip laterally by 5 mm, to prevent collision with the posterior aspect of the acetabulum. The range that could be tested was limited by the field of view of the ultrasound probe. The range tested was from 5 mm medial to 7.5 mm lateral displacement, to ensure the maximum cross section of the femoral head remained within the image capture volume.

4.4.3.3. Repeatability of image analysis

The repeatability of the image analysis was investigated for the LHC index by selecting 14 images at random from the side-lying image set. The assessors were blinded to the level of translation of the hip and the relative orientation of the segments. They were not blinded to whether the image volumes were taken from the left or right hip phantoms, as the assessors needed to be able to correctly orientate the image for analysis. Three assessors (myself (RK) and MJ, LJ) analysed the images according to the methods described in section 4.4.2. RK who had developed the indices was familiar with the content of the image volumes. The other two assessors (MJ and LJ) had minimal experience in analysing 3D ultrasound data. RK ran a 1-hour training session, followed by supervised analysis of 10 training images. Each assessor measured the LHC for each of the test image volumes on three separate sessions. The image order was altered for each of the sessions and the repeat session took place two weeks after the first. The values for the indices were sent to RK for analysis. The repeatability of the image acquisition was not investigated as the rig set up was designed to minimise variability at acquisition, therefore any study of the repeatability of the image acquisition would not be a true reflection of the variability of clinical image volume acquisition.

4.5. Data analysis

We hypothesised that LHC and FHPPR would be independent of variation in relative orientation of the femoral and pelvic hip segments, within +/- 20 degrees of ‘neutral’ in each anatomical axis (internal/external rotation, ab/adduction, flexion/extension).

Simple linear regression was used to test the dependence of FHPPR and LHC on the relative orientations of the femoral and pelvic segments in each of the 3 axes of rotation (flexion-extension, abduction-adduction and internal-external rotation).

We hypothesised that LHC would be directly proportional to medio-lateral translation of the femoral head relative to the pelvic segment but would be insensitive to posterior translations. Similarly, we expected FHPPR to be directly proportional to posterior displacement of the femoral segment relative to the pelvic segment but insensitive to lateral translations. Linear regression was used to investigate the dependence of FHPPR and LHC on each of medio-lateral and antero-posterior translation of the femoral segment relative to the pelvic segment.

Intra assessor variation was calculated by computing the variance within each assessor and within each image. For each assessor, the average variation across the 14 image sets was taken as that assessor's intra-assessor variation. To compute inter-assessor variance, the total variance in all sessions and all assessors was computed for each of the 14 images. This variance represents the total variance, i.e. inter plus intra assessor variance. The average intra-assessor variance was deducted from this total variance to generate the inter-assessor variance. The standard deviations for each of intra-, inter- and total- variance were then computed.

4.6. Results

4.6.1. Sensitivity of the lateral head coverage index to rotation

Axis of rotation	*Hip (status)*	*ΔLHC/°*	*R^2*
Flexion/Extension	Right (as printed)	-0.002	0.647
	Left (+ 40° ext rot)	-0.001	0.579
Ab/Adduction	Right (as printed)	0.005	0.871
	Left (+ 40° ext rot)	0.005	0.926
Internal/External rotation	Right (as printed)	-0.001	0.787
	Left (as printed)	-0.002	0.953

Table 11: Relationship between LHC and relative anatomical rotations of the femoral segment relative to the pelvic segment.

Figure 22 illustrates the sensitivity of LHC to rotations about each of the 3 orthogonal anatomical axes. For each of the rotations the relationships with LHC were linear. The strongest relationship, which also had the steepest gradient, was the relationship about the ab/adduction axis. For each degree of adduction the LHC reduced by 0.5%

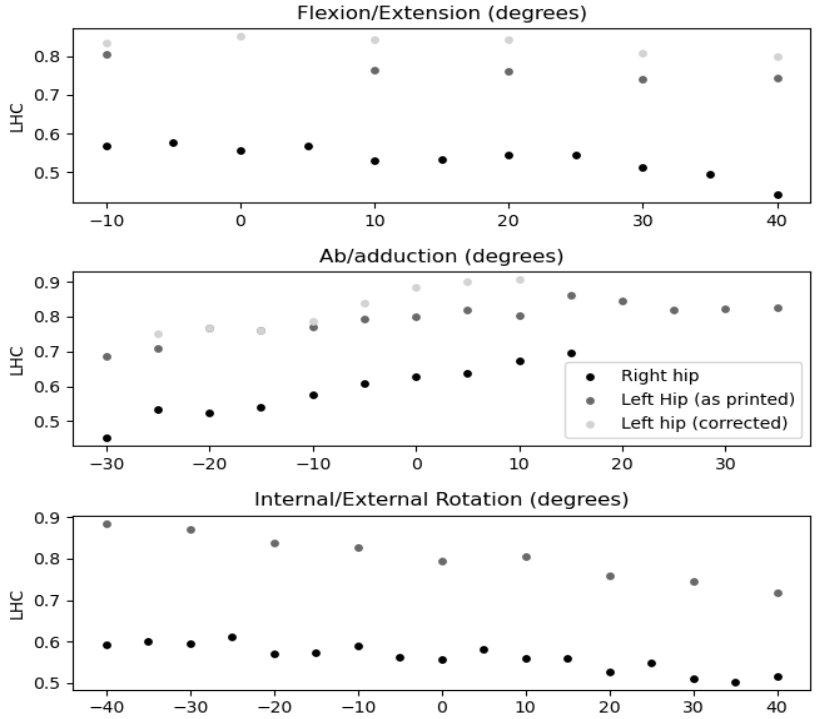


Figure22: Subplots showing the relationship between anatomical rotations and LHC for both the left and right hip phantoms. Note: images were captured at increments of 5degrees for the right hip, a reduced data set were collected for the left hip. This decision was taken to reduce the analysis burden whilst still confirming a consistent relationship between the LHC and the anatomical rotations. The left hip were captured in both printed and 'corrected positions' to ensure that starting position did not impact the relationship between rotations and index. When assessing rotation, it was not possible to start with the left hip in the corrected position as there was not enough range in the rig to then assess a further 40 degrees rotation.

(R^2 = 0.871 and 0.926 for the right and left hips respectively). To reduce the number of images to be analysed, the right hip was imaged at all points (increments of 5°), whilst the left hip was imaged in increments of 10°. The left hip was imaged in both the printed and corrected positions. This was deemed acceptable as the relationships between the rotations and the indices were consistent regardless of the chosen hip, thus more granular data was not deemed required.

LHC was less sensitive to rotations about both the flexion/extension axis and the internal/external axis. There was a 0.1-0.2% per degree change with moderate to strong correlations, Table 11 (Appendix 2 for full results). LHC reduced with internal rotation and flexion, i.e. reducing the coverage of the femoral head by the acetabulum.

Axis of rotation	*Hip (status)*	*ΔFHPPR/°*	*R^2*
Flexion/Extension	Right (as printed)	0.001	0.94
	Left (+ 40° ext rot)	0.002	0.92
Ab/Adduction	Right (as printed)	-0.000033	0.005
	Left (+ 40° ext rot)	-0.001	0.701
Internal/External rotation	Right (as printed)	0.001 (*0.000)	0.119 (*0.520)
	Left (as printed)	0.001	0.633

*Table 12: Relationship between FHPPR and relative anatomical rotations of the femoral segment relative to the pelvic segment. *indicates relationship once the outlier data point is removed.*

4.6.2. Sensitivity of the femoral head posterior position ratio to rotation

FHPPR was less sensitive to anatomical rotations than LHC, Table 12 and Figure 23. The only axis where a strong correlation was observed was the flexion/extension axis where for both hips there was a 0.1-0.2% change in FHPPR for each degree change. As the hip became more flexed the FHPPR increased this was due to the distance between the anterior border of the acetabulum and the femoral head increasing in the image volumes. The left hip was imaged in the corrected position only, this was due to the greater trochanter shadowing the femoral head in the printed, highly rotated,

position. As with LHC, the right hip was imaged in all positions, whilst the left hip was imaged in a reduced number of positions.

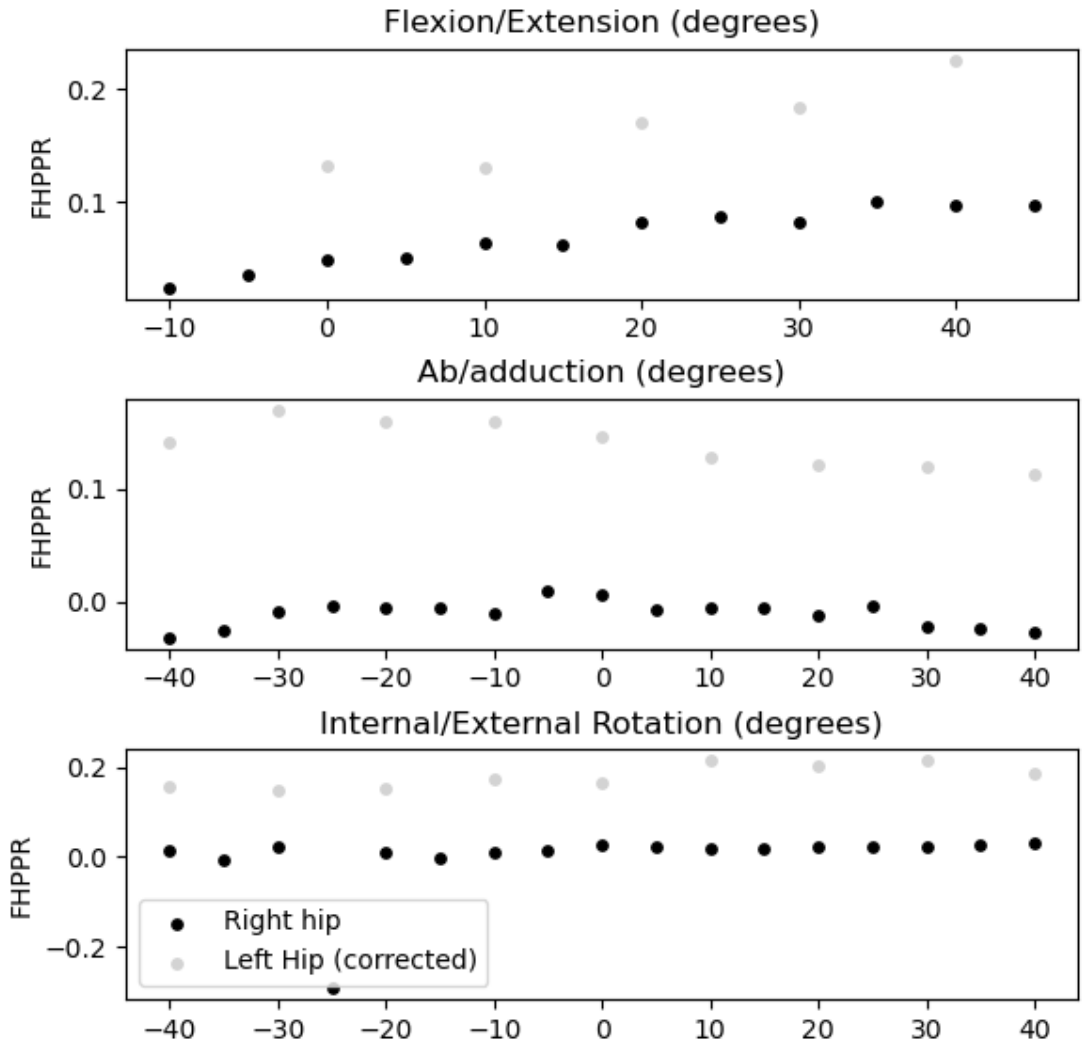


Figure 23: Subplots showing the relationship between anatomical rotations and FHPPR for both the left and right hip phantoms. Note: images were captured at increments of 5degrees for the right hip, a reduced data set were collected for the left hip. This decision was taken to reduce the analysis burden whilst still confirming a consistent relationship between the FHPPR and the anatomical rotations. The left hip was imaged in the corrected position only as in the original position the greater trochanter shadowed the femoral head rendering it impossible to analyse the images.

4.6.3. The sensitivity of lateral head coverage to translational displacements.

LHC had a very strong linear relationship with medio-lateral displacement with R^2 coefficients of 0.99 for both the left and the right hip (Figure 24). Simple linear

regression for the left hip resulted in a gradient of the slope of -0.028 (95% CI: -0.029 to -0.027). For the right hip the linear regression showed similar results, the gradient was -0.027 (95% CI: -0.028 to -0.026) (Appendix 3).Practically these results mean for a 1 mm lateral displacement there was approximately 3% change in LHC. LHC was also sensitive to out-of-plane translations of the FH. It had a similar strength relationship to posterior displacement as to lateral displacement.

It should be noted that to be able to posteriorly displace the femoral heads without obstruction from the posterior acetabular wall, the femoral heads were laterally displaced by 5mm and then posteriorly displaced in increments of 1mm.

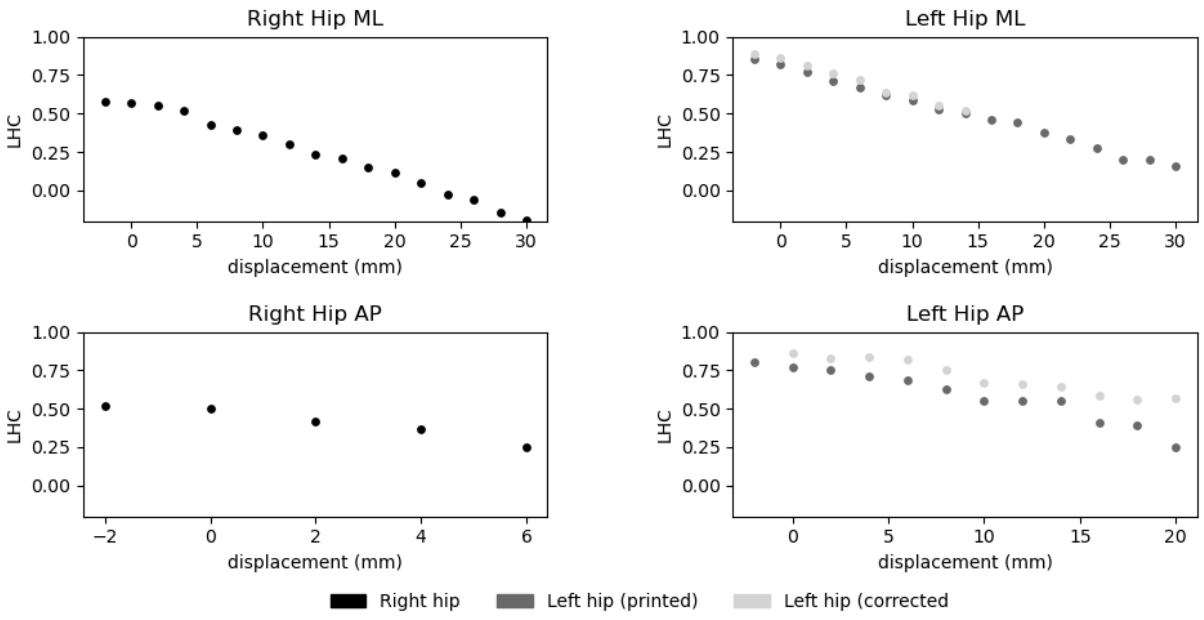


Figure 24: Relationship between LHC and lateral displacement and LHC and posterior displacement for both phantoms.

4.6.4. The sensitivity of femoral head posterior position ratio to translation displacements

FHPPR had a strong linear relationship with posterior displacement (R^2= 0.98/99), with a gradient of 0.025/mm and 0.024/mm for the left and right hips respectively (Figure 25). These results indicate that a 1mm posterior displacement increased the distance between the anterior border of the acetabulum and the femoral head centre by approximately 2.5% of the femoral head diameter for both the left and the right phantoms. FHPPR was found to be largely insensitive to lateral displacement (right hip R^2 = 0.39, gradient 0.011 (95% CI = 0.004 – 0.018), left hip R^2 = 0.66 gradient 0.004 (95% CI = 0.002 – 0.005)), see Appendix 3 for full statistics.

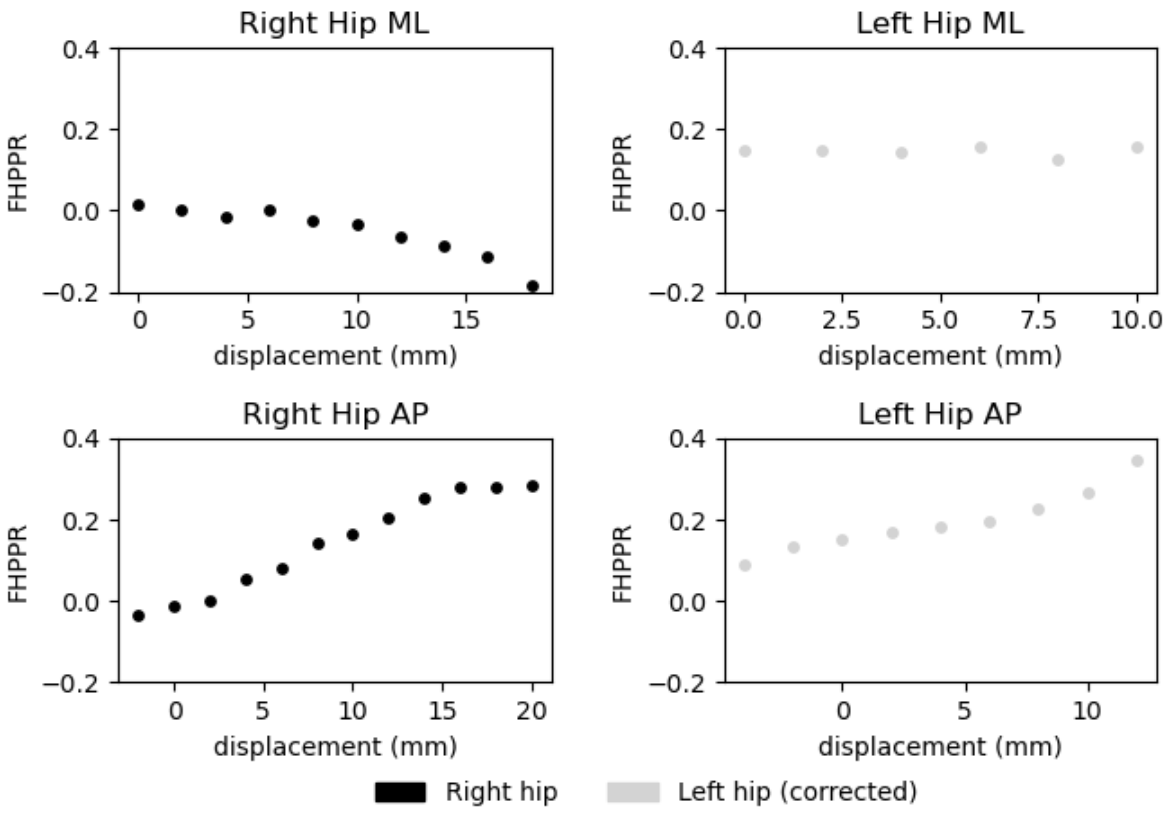


Figure25: Relationship between FHPPR and posterior displacement and FHPPR and lateral displacement for both phantoms.

4.6.5. Repeatability

Table 13 shows the results from the intra- and inter- assessor variation as well as the total variance in the measurement of LHC**.** The standard deviation of the within assessor error was under 3% whilst the standard deviation of the between assessor error was 2.44%. The overall error was approximately 4% LHC.

LHC	*Intra assessor error*	*Inter assessor error*	*Intra + inter assessor error*
Variance	8.74% (4.13% - 11.7%)	5.95%	14.69%
Standard deviation	2.89% (2.03% – 3.42%)	2.44%	3.83%

Table 13: LHC intra, inter and total repeatability data across 14 images, 3 assessor and 3 sessions.

Figure 26 is a boxplot illustrating the variation in measurement of LHC across each of the individual images. Image 9 shows the greatest variation in measurement and was also the most uncovered hip. LHC performed well particularly if the hip was well positioned (i.e. greater than 70% covered).

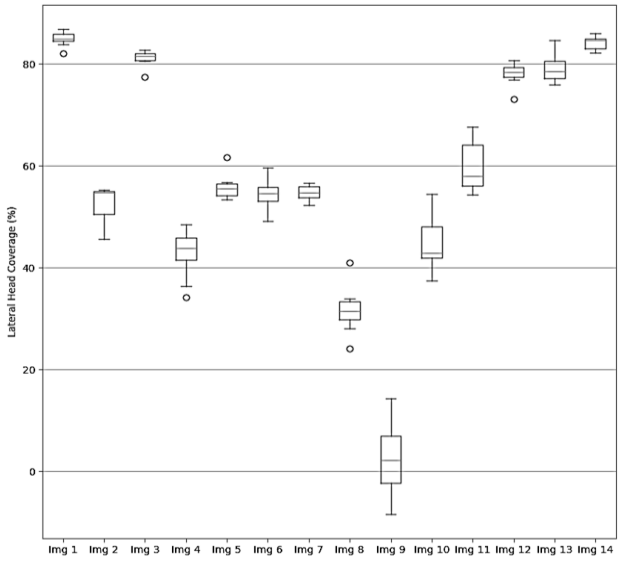


Figure 26: A box plot showing the total variation between LHC measures for each of the different images – across assessor and session. (N = 9 for each image).

4.7. Discussion

The objective of the investigations reported in this Chapter was to evaluate the sensitivity of ultrasonic indices of lateral and posterior hip displacement to in- and out-of-plane displacements and to rotation using anatomically realistic phantoms created using additive manufacturing. Experiments were repeated using 3D prints of two different hips. It was hypothesised that LHC and FHPPR would be independent of variation in the relative orientation of the femoral and pelvic segments within an envelope of neutral (+/- 20 degrees of neutral). The envelope was selected as the

outer limits of what is easily detected by the eye of the clinician, i.e. if it was not possible to manipulate the child limbs into a neutral, or near neutral position the clinician would be able to report it without the requirement of a measuring tool.

Further it was hypothesised that LHC would be directly proportional to medio-lateral translation of the femoral head relative to the pelvic segment but would be insensitive to posterior translations. Similarly, FHPPR would be directly proportional to translations of the femoral head relative to the pelvic segment in the antero-posterior plane and insensitive to translations in the medio-lateral plane.

Finally it was hypothesised that the repeatability of image analysis would be similar or better than that reported in the literature for the measurement of RMP. Measures of repeatability of RMP vary in the literature, however the standard deviation of the variance is usually quoted as being between 3-4%.

4.7.1. True lateral and posterior displacements

The results showed the indices were highly sensitive to displacements within the plane of measurement, i.e. LHC was highly sensitive to coronal plane displacements or lateral displacements and FHPPR was highly sensitive to sagittal plane displacements or posterior displacements. These results were anticipated, and the experimental results strongly support the hypotheses. The correlation coefficients were greater than 0.98, indicating a near perfect linear relationship. LHC describes true lateral displacement (R^2 ~0.99) however it is also highly sensitive to the posterior displacement (R^2 = 0.89 for the right side and R^2 = 0.92 for the left side) – with near identical gradients (right side: -0.025 (95% CI: -0.031 to -0.02) and left side: -0.028 (95% CI: -0.032 to -0.025)). Some sensitivity was expected due to the construction of

the LHC measurement which used the lateral aspect of the acetabulum border in the slice selected to have the maximal cross-sectional area of the femoral head. The acetabulum retracts posterior-medially and hence has a similar appearance to a laterally displaced femoral head. Further, to be able to posteriorly displace the femoral head in the experimental set-up, the femoral head was laterally displaced by 5 mm and then posteriorly displaced in increments of 1 mm. In the regression, to estimate the relationship between LHC and posterior displacement, the confidence intervals were wider and the coefficient of determination lower, indicating the linear model did not fit the data as well. The confidence intervals were also affected by a lower sample size in the regression model for the right hip. This relationship between LHC and posterior displacement of the femoral head relative to the acetabulum does have clinical implication. Used as a solo measurement LHC can detect displacement of the hip, however the plane of the displacement cannot be inferred from the isolated measurement of LHC. However, used in conjunction with FHPPR the direction of the displacement can be determined and likely the clinical utility of LHC is improved. In this study we were not able to validate the use of FHPPR in the clinical population against an established 3D imaging modality, such as CT however the potential clinical utility in conjunction with a measure of lateral displacement, either LHC or RMP from X-ray is explored in chapter 6.

Further investigation of image volumes, where the LHC measurements were similar but the hip positions were different, was conducted to identify markers to indicate when an LHC measurement was measuring a true lateral displacement or the measurement was confounded by posterior displacement. Two set-ups where identified, in the first set-up the femoral head was laterally displaced by 12mm and in the second set-up the femoral had was posteriorly displaced by 8mm. Figure 27 shows the rendered images

constructed from the two image volumes. Figure 27 (i & ii) shows the coronal plane slice of the two ultrasound volumes that was selected for analysis, i.e. the slice of maximal cross sectional area of the femoral head (in both the coronal and sagittal planes). The LHC in both images was measured to be 0.1. Figure 27(iii & iv) show the coronal plane view of the rendered image volumes, the red render is the surface of the femoral head that was visible in the ultrasound volume, the yellow render is the visible pelvic segment, the acetabular border is included in this render. A green sphere, representing the estimated best fit sphere, is also displayed as an estimate of the femoral head. A clear difference in the position of the femoral head relative to the acetabular boarder can be seen in the images. Figure 27 (v&vi) shows the same rendered images but viewed in the transverse plane, i.e. looking down the body. The white dotted line shows the image slice that was chosen for analysis. From this view it can be seen that as you trace the acetabular border posteriorly it retracts medially, and hence for a posteriorly displaced hip, where the slice of maximal femoral head cross sectional area is chosen for analysis, appears to have reduced lateral coverage of femoral head by the acetabulum when measured by LHC. Due to the construction of the X-ray images, a posteriorly displaced hip, which might have reduced acetabular coverage in its position, will not be detected.

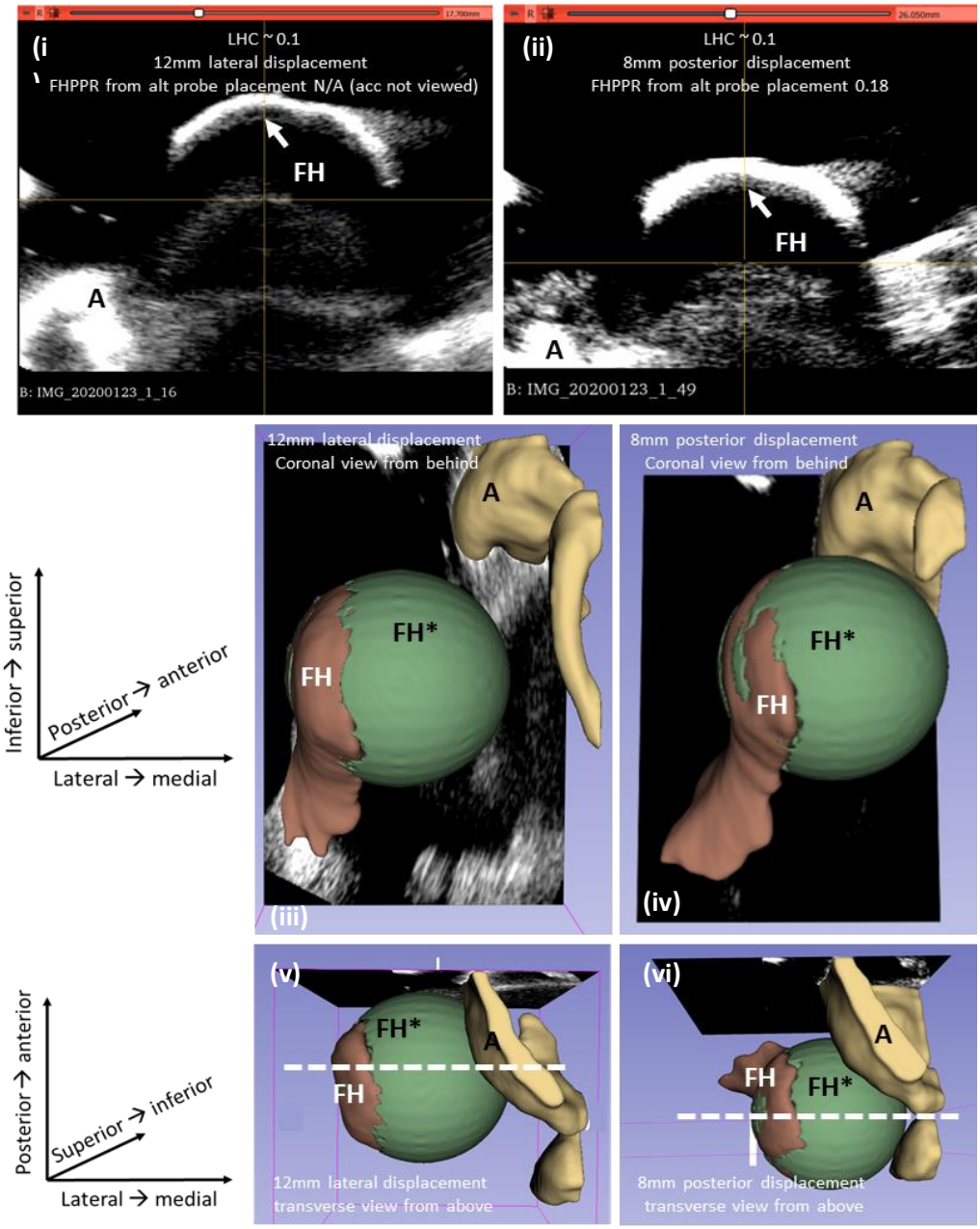


*Figure 27: Further exploration of two image volumes where the LHC measurement was similar. (i & ii) - the ultrasound slices selected for LHC measurement. (iii & iv) - the coronal views (from behind) of the rendered hips. (v & vi) – transverse view of the rendered hip, with the position of the slice selected for analysis marked highlighted. A = accetabulum render, FH = femoral head render, *FH = sperical estimate of femoral head.*

Figure 28 shows a series of coronal plane slices, around the slice selected for analysis, for each of the image volumes that were investigated further. Each slice is 2mm spaced from the next slice, the 3rd slice down in both columns is the slice chosen for

analysis. It can be seen that for the image volume where the femoral segment is displaced laterally, the acetabular border will remain relatively static in the frames either side of the frame chosen for analysis (i.e. the frame of maximal cross sectional area of the femoral head). However, when the femoral head is displaced posteriorly the acetabular border appears similar in the frame of maximum cross-sectional area, but its position alters significantly in the surrounding frames. This is due to the shape of the acetabulum. At the most lateral point the acetabular border is broad, however when moving this image around posteriorly the acetabulum begins to retract medially.

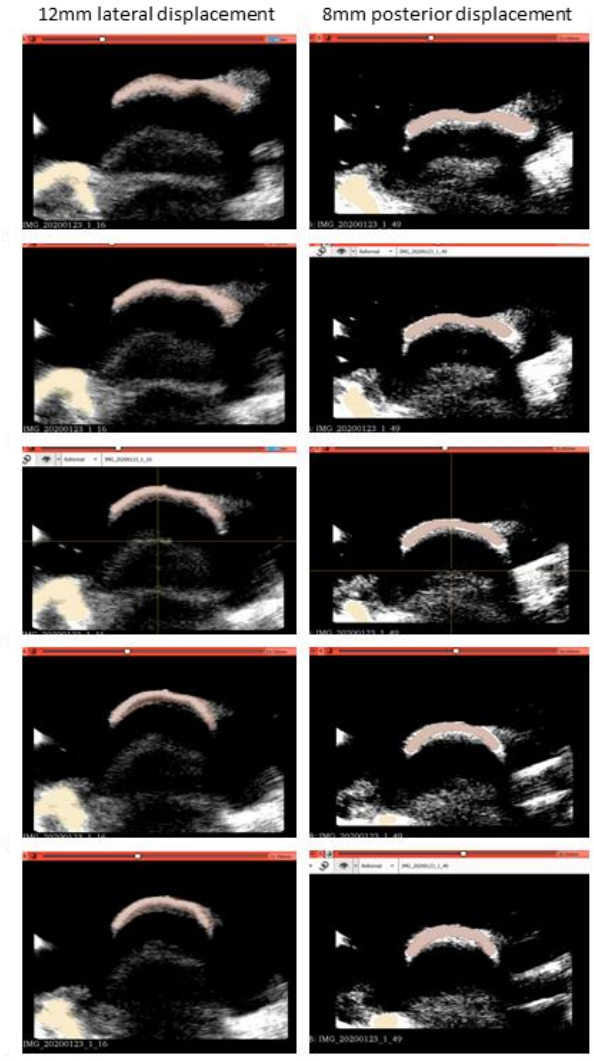


Figure 28: Coronal plane slices, taken at 2mm intervals for each of the phantom image volumes. The femoral heads have been highlighted in red and the acetabulum has been highlighted in yellow.

In comparison to LHC, FHPPR was shown to be much less sensitive to out of plane displacements. The regression results for the left hip showed the sensitivity to lateral displacements to be an order of magnitude lower in comparison to the sensitivity to posterior displacement, with moderate correlation (R^2 = 0.663). The right hip showed a weaker correlation (R^2 = 0.39) but the regression coefficient showed that it was more sensitive to lateral displacements in comparison to the left hip. Given the insensitivity of FHPPR to lateral displacement, if used in conjunction with LHC it could be used to differentiate between situations where posterior displacement is confounding the LHC measurement and true lateral displacement.

4.7.2. Anatomical rotations

We hypothesised that both indices, LHC and FHPPR, would be insensitive the anatomical rotations, within ±20° of neutral. This hypothesis is only partially supported by our findings. Our methodical assessment of both indices to femoral rotations about each of the three anatomical axes revealed that both LHC, and to a lesser extent FHPPR, were sensitive to the rotation in the plane of measurement. In the case of LHC, the coronal plane rotation is ab/adduction, our results suggest that LHC has a systematic dependence on ab/adduction, with a change of 1° of ab/adduction resulting in a change of 0.5% (CI: 0.3%-0.6%) in LHC. This linear relationship had very strong correlations observed for the measurements taken from both the right and left hip images (R^2 = 0.871 and 0.926 for the right and left hips respectively). In the case of FHPPR, the rotation within the plane of measurement is flexion/extension. We observed a similar relationship between FHPPR and flexion/extension, with approximately 0.2% (CI: 0.1% - 0.4%) change in FHPPR for every 1° change in

flexion/extension. The other rotations displayed weaker and less sensitive relationships.

To identify the cause of the relationship between LHC and ab/adduction that was observed the image series from the right hip was explored further. It was observed that as the hip moves from adduction to abduction, the proportion of the femoral head that can be seen in the coronal slice reduces. To ensure that this was a genuine phenomenon, much like that observed by Reimer in his investigations[27], the images were evaluated further. The effect of reducing the surface area of the estimate of femoral head diameter (FHD) was investigated. The largest FHD was recorded at the neutral position and the variability in FHD was minimal in comparison to the variation in LHC, confirming that the relationship between LHC and ab/adduction is a genuine dependence and not a conflated by a systematic measurement error. Figure 29 shows a series of images taken from 30° adduction through to 10° abduction. It was observed that the most lateral surface of the femoral head appears to be more lateral, relative to the lateral border of the acetabulum, when the hip is adducted in comparison to the image where the hip is abducted. This is in keeping with Reimer's observation, Figure

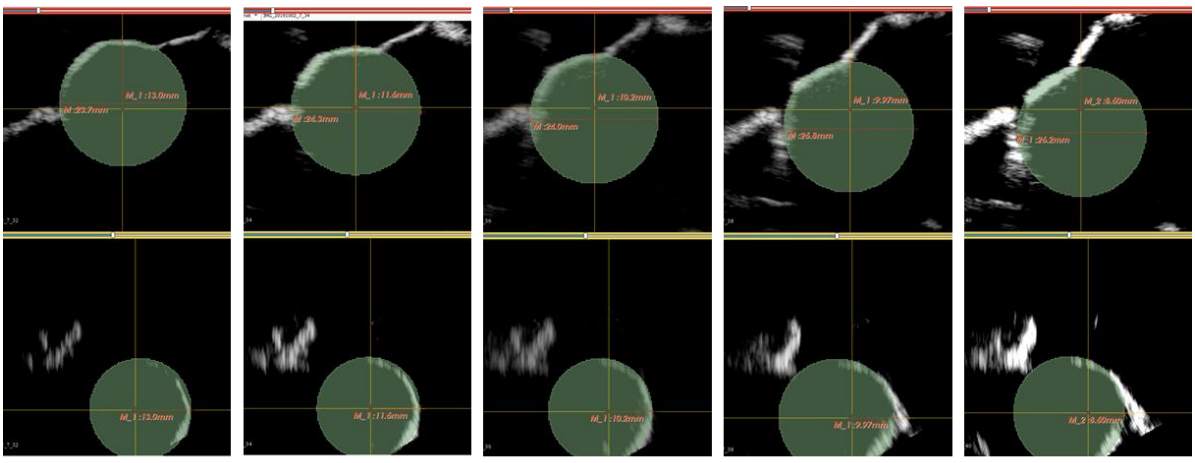


Figure 29: A series of images showing the slice selected for analysis as the right hip is moved from 30 degrees adduction to 10 degrees abduction. The top images are the coronal slice and the lower images show the sagittal plane slices used to identify the slice selected for analysis.

30 shows the X-ray images published in Reimer's thesis demonstrating this phenomenon[27] (Figure 30) also illustrated how internal rotation can affects the projected image of the hip in X-ray. In our experiments we observed a 1-2% change in LHC for each 1° change in rotation.[27]

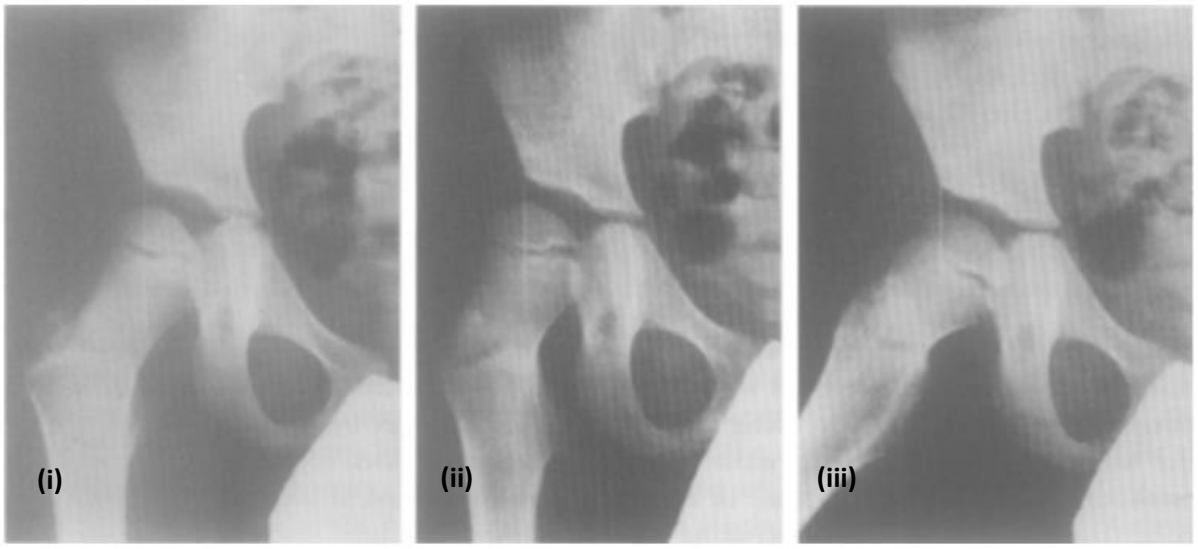


Figure 30: Images taken from Reimer's thesis[27] showing the effect of rotation and abduction on the femoral head positioning. (i) maximum internal rotation, (ii) neutral positioning, (iii) abduction[26]

4.7.3. Repeatability study

Our repeatability study results showed the standard deviation of the combined error in the LHC measurement to be below 3.9%. Cliffe *et al*[75] reported a standard deviation of 3.8% RMP across assessors and sessions, their results are very similar to others[76–78].

Figure 31 displays a box and whisker plot of each of the variations in LHC measurements for each of the 14 images included in the repeatability study. Visual inspection indicated that image 9 had the greatest levels of variance. This image is also the most extreme LHC measurement with a median LHC measurement of less than 5% coverage. Exploring the variation in image 9 further, it can be concluded that

the variation is primarily due to the measurement of femoral head diameter (Figure 31).

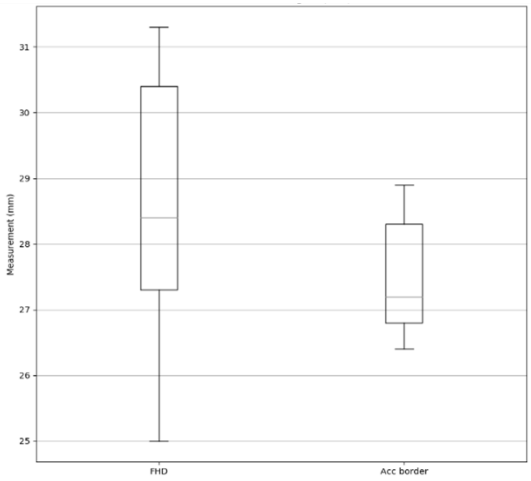


Figure 31: A boxplot showing the variation in measurements for both femoral head diameter and acetabulum border for image 9 across the 3 assessors and 3 sessions.

The cause of the variation in femoral head estimate was investigated by analysing the slice selected for analysis by each assessor. Figure 32 shows the slice selected for each image across assessors and sessions. Image 7 has the greatest range of selected slices, with the extremes at 6.9mm into the image volume and 9.9mm into the volume (a 3mm range).

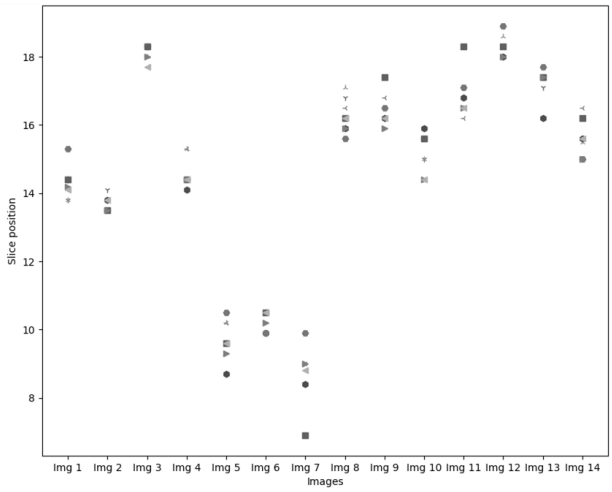


Figure 32: Slice selection for each image, across assessors and sessions

Modelling this extreme scenario, assuming that the maximum cross-sectional area was at the slice at 9.9mm and the FHD was measured at 38.9mm (the average FHD across both hips and all measurements). The estimated FHD measured in the slice taken at 6.9mm can be calculated (Figure 33, Equation 10). Assume *x* = 3mm and *r* = 19.45mm, FHD_1 = 38.4mm. This equates to a 0.5mm difference in estimated FHD at the most extreme example in the repeatability dataset, hence it can be concluded that slice selection has minimal impact on FHD estimate.

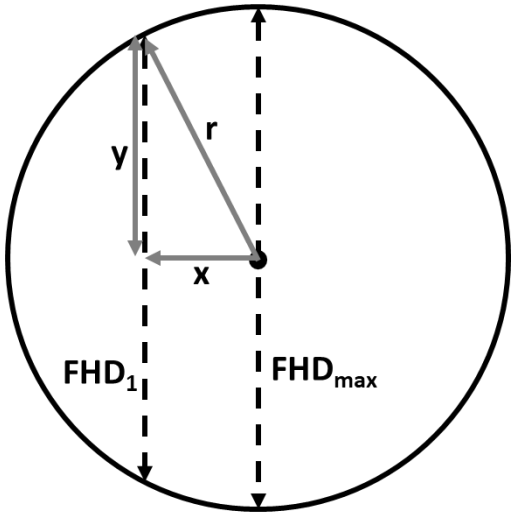


Figure 33: Diagram showing the relationship between different slice selection and estimated FHD

$$FHD_1 = 2y = 2\left(\sqrt{r^2 - x^2}\right)$$

Equation 10: Calculation for FHD from different slice selection

The FHD is made from best fitting a sphere to the lateral or anterior curvature of the femoral head. It is likely that the contrast and brightness (both variable at the analysis phase) have a greater impact on the identification and sphere fitting. In these

experiments the ultrasound parameters at acquisition (namely gain and depth) were kept constant throughout the imaging. However, clinically these factors may influence the identification of bony surfaces in the image analysis.

4.8. Clinical implications

These *in vitro* experiments were designed to assess, as best as possible, the robustness of the new indices. By restricting sources of variance and systematically varying a single element the ground truth capabilities of the measurements were assessed. Overall, the indices proved themselves to be reasonably robust and as such their clinical utility should be evaluated further in an *in vivo* scenario. However, there are some results which warrant further exploration and consideration regarding their clinical impact.

Our results showed that for every 1° increase in adduction LHC reduces by 0.5%. Practically this means that for the range ±20° about neutral there would be a difference of 20% in LHC, which would be a clinically significant change. Defining a clinically significant change is challenging but it is possible to relate the change in LHC or FHPPR to a true change in displacement using the translational experimentation results. The translational experiment revealed that 1mm of genuine lateral displacement resulted in a reduction of LHC by approximately 3% - this would equate to a 6° shift in adduction. If LHC were to be used clinically it would be necessary to ensure that a neutral position in the ab/adduction axis were adopted and in cases where this is not possible/comfortable the level of adduction could be measured so that it can be accounted for in the LHC calculation, with the limitation that the clinical measurement itself may be subject to error. Further experimentation, over a wide range of hips, would be needed to define the ‘correction’ factor. Once defined the

equation would then require validation. This would require different hips to be imaged to those used in the defining of the equations.

Hip adduction contractures, or windswept hip deformities are present in approximately one third of individuals classified at GMFCS level III –V[74]. The level of deformity varies[74,110,111], but can still often be corrected with the application of pressure. RMP, measured from planar X-ray is also sensitive to windswept positioning as demonstrated in Figure 30[27] above, unlike X-ray ultrasound assessment offers the flexibility of manipulating and holding the limb during image acquisition and repeating images when positioning is not considered optimal.

The other anatomical axes showed regression coefficients that equated to either 0.1% or 0.2% change in LHC per degree of rotation around each of the axes. This means that to keep the change in LHC to within 3% (which is an arbitrary change but equivalent to 1mm of genuine displacement and in line with intra and inter assessor repeatability results), the rotation would need to be within 15°-30° of neutral. Clinically these are levels of rotations that can either be identified and measured, or in most cases accommodated.

FHPPR was less sensitive to anatomical rotations in comparison to LHC. For all rotations, except flexion/extension of the left hip, a 10° rotation equated to a 1% change in FHPPR. When this result was compared to the translational results, where a 1mm genuine posterior displacement resulted in a 2.5% change in FHPPR, it can be concluded that within ±10° neutral, the change in FHPPR would be insignificant, and likely hard to detect.

The translational experiments were conducted for two reasons, to assess the indices' ability to measure true displacements in the plane of the measurement and to quantify

the indices' dependence on out of plane displacements, specifically those in the orthogonal plane. Our results suggest that FHPPR is relatively insensitive to out of plane displacements and highly sensitive to displacements in the intended plane of measurement, a result which would support the further evaluation of the index as a clinical tool. However, LHC was shown to be highly sensitive to both in plane and out of plane displacements. Clinically, it is still useful to understand if there is a displacement, and these results suggest that LHC is highly sensitive to detecting displacements generally. However, understanding the direction of displacement is clinically very valuable when evaluating treatment options. Currently, a primary weakness of the measurement of RMP is the relative insensitivity to out of plane displacements, LHC demonstrated sensitivity to displacement irrespective of the plane. It is likely that the clinical utility of 3D ultrasound in the assessment of hip displacement is in the combined information from both the LHC and FHPPR. Allowing the level of displacement and the direction of the displacement to be quantified. However, the LHC index, was designed to detect displacement in a similar way to RMP, using similar landmarks but reconstructed from a different imaging mode so not an entirely equivalent index. It is therefore of clinical interest to conduct a levels of agreement study to assess the performance of LHC in comparison to RMP in the detection of lateral displacement.

4.9. Study limitations

4.9.1. Missing data

It was not always possible to analyse the images; however it was not felt that the missing data impacted the results reported as in all cases the missing data occurred at the extreme positions of image collection. Additionally, for the translational experiments, the missing data only affected the results from experiments in the orthogonal plane to the measured index. In some positions the femoral head was shadowed by the femoral neck or greater trochanter. In these situations, the images could not be analysed. This applied to both hips in the ab/adduction tests performed in the side lying position. In other images, the maximal cross section of the femoral head was outside the image volume, thus the FHD could not be estimated. This situation arose for both hips in the supine position when the femoral head was displaced laterally, i.e. when assessing FHPPR sensitivity to medial-lateral displacement. Finally, the acetabular border could not be clearly identified in the posterior displacement test for the right hip in the side lying position from 7mm to 15mm posterior displacement as it appeared to become shadowed by the femoral segment.

4.9.2. Between session errors

The rotational and translational experiments were conducted on separate days, despite best efforts to standardise the start positions, the tolerance on the printed guides was too great resulting in the inability to reliably replicate the CT scan positioning between sessions. For this reason, the start positions differed between sessions, however within a session, the start position was regained at multiple points through each test and an image taken. Our repeated neutral images showed a

standard deviation in measurements of approximately 2.5% and 2% for LHC and FHPPR respectively. These errors incorporate variance within a single assessor's analysis and variance in the rig's ability to regain a neutral position. The LHC error is comparable to our LHC intra assessor error from the repeatability study. We therefore can conclude, with confidence, that the rig performed reliably and was a negligible source of error.

4.9.3. Anatomical accuracy of the phantoms

There are two key limitations regarding anatomical fidelity with the approach used in this *in vitro* study. The first is simplicity of the models, which lack soft tissue structures. It is possible to create tissue mimicking ultrasound phantoms, these phantoms include layers of different materials deigned to simulate the acoustic properties of bone, muscle, fat etc. Whilst the ideal phantom would have included different materials to mimic different tissue properties, the primary goal was to be able to simulate many different positions of the femoral and pelvic bony segments of the hip. For practical reasons it was not possible to construct a phantom with the positional flexibility required that was also contained in a tissue mimicking casing. Instead the bony segments were printed in a plastic with acoustic properties that were reasonably representative of bone[112].

The second is the lack of variety in the bony geometries modelled. Two hips were printed, both created from a CT scan of a single individual with CP. Heterogeneity within CP has been widely discussed and is known to impact and challenge our understanding of the true underlying uncertainty in an image. For example, the more displaced a hip is, typically the more rounded the acetabular border becomes, increasing the challenge of reliably identifying the lateral edge of the acetabulum, the

landmark used to draw Perkins line. From this study it is impossible to understand the impact that anatomical differences between individuals with CP would have on the measurements assessed. The results from this *in vitro* set up were expected to represent 'the best one can expect' from the indices.

4.10. Conclusion

The *in vitro* set-up performed excellently, facilitating the development and testing of both coronal and sagittal plane indices of hip displacement. Both indices were observed to robustly quantify displacement in the plane they were designed to measure. LHC showed sensitivity to out of plane displacements. These sensitivities were not observed for the FHPPR index. Both indices showed some sensitivity to rotations about the anatomical axis in the plane of the image used for analysis. The greatest sensitivity observed was the impact of ab/adduction on LHC, however this sensitivity has been observed in X-ray measurement of lateral displacement (RMP) and is due to the centre of rotation of the femoral head in that axis. Further studies using a greater range of hip geometries are needed to support these findings.

To better understand the clinical use of these indices an *in vivo* study is required, firstly to compare to the gold standard measure of lateral hip displacement, RMP, and secondly to compare to conventional modality of 3D imaging the hip such as CT or MRI. This later study is particularly important in the validation of FHPPR for clinical use.

5. A preliminary validation of 3D ultrasound in assessment of lateral hip displacement: A comparison to the current clinical standard

5.1. Overview

This chapter can be considered as a preliminary concurrent validation of 3D ultrasound in the evaluation of lateral displacement of the hip in children with cerebral palsy. The limitations of RMP have been discussed at length in the preceding chapters, it therefore may seem illogical to compare a new method with a flawed one. However, RMP remains the most prevalent measurement of hip displacement in this population and therefore can be considered the clinical standard, against which the performance our index of hip migration could be compared. Concurrent validation is a measure of the agreement between a new index or test with an established measure. This chapter will focus on quantifying the level of agreement between LHC, derived from 3D ultrasound images and RMP derived from 2D radiographs.

Ethics approval was sought from a research ethics committee to enrol up to 40 children with cerebral palsy undergoing routine hip surveillance X-rays at the Evelina Children's Hospital between 2017-2020. These children were asked to attend an ultrasound appointment within 2 months of their surveillance X-ray. The RMP measured on the X-rays was then compared to the LHC measured from the 3D ultrasound volumes collected. The results show a strong correlation between RMP and LHC, with similar levels of intra- and inter- assessor reliability in the analysis of LHC, compared with those reported for RMP[76–78].

5.2. Preliminaries

The Health Research Authority (HRA) approval and national ethics approval from Wales REC 7 committee (study number 17/WA/0093) was granted for this study (Appendix 4). Local approvals from the research and development team and the clinical research facility (CRF) were granted. This allowed the study to be conducted at the CRF at St Thomas' Hospital, a dedicated unit specifically for research purposes.

5.2.1. Sample size justification

As the use of 3D ultrasound to assess the hip in older children with cerebral palsy was novel, there was limited pilot data available to inform the sample size calculations. A simulation, much like the one described in chapter 3, was developed to define the sample size for the study. The simulation was constructed to understand the chance of a type 1 (false positive) and type 2 (false negative) error. It was decided that the required sample size would be the minimum number of subjects where the type 1 and type 2 error were less than 5%.

It was assumed that Bland Altman statistics would be used to assess the level of similarity between the RMP and the LHC measurements. The following input assumptions were made, these assumptions were considered conservative to ensure the sample size was not underestimated:

1. The data would be normally distributed.
2. The mean true RMP across the data would be RMP=20%.
3. The standard deviation of the data would be RMP= 10%.
4. The standard error in x-ray technique is 15% - estimated form ranges quoted in the literature.

5. The standard error in the new measurement technique (3D ultrasound) is 5% - estimated from previous muscle volume studies.
6. No bias between the systems.

In a similar fashion to the Monte Carlo simulation described in detail in chapter 3, a 'true' data set was created using randomly generated normally distributed data defined by the mean and standard deviation from the initial assumptions. The 'measured' data points of this true data point were then created for each of the two techniques. The difference between the two measured data points was calculated. This was then repeated for different sample sizes, i.e. initially 10 true points (obeying the initial assumptions) with associated 'measured' points were generated. This was then repeated and the average difference between the techniques, 95% confidence limits, and bias limits were calculated for this data set. The exercise was then repeated for different sample sizes. For each situation, the number of times a bias was detected, when there was no true bias, was counted. This is an estimate of the type 1 error.

The same simulation was then run on the same situations however assumption number 6 was changed. A true bias of 10% was added to the model. Once all the situations had been simulated the number of times a bias of 10% or greater was detected was counted. This is an estimate of the type 2 error.

The sample size was increased in increments of 10 to understand the impact of sample size on the type 1 and type 2 error rates for the situation modelled. From these results, it was concluded that at least 40 hips would be required to adequately power the study. Any hips which had received surgery to modify the shape of the acetabulum were excluded from the study, for this reason not every recruit would have both hips

included in the study. Ethical approval was granted for a maximum of 40 individuals, aged between 2 and 16 years, to be recruited to the study.

5.3. Objectives

The primary objective of this study was to perform a preliminary concurrent validation of 3D ultrasound for assessing lateral displacement of the femoral head in children with cerebral palsy. For this, a new ultrasound-based index, LHC (defined in chapter 4) was compared to RMP. The secondary objective was to establish the intra and inter-assessor reliability of our new index, LHC.

The clinical study had three phases, two are discussed here and the final phase is discussed in Chapter 6.

1. Comparison between RMP and LHC: we hypothesised that LHC would be highly correlated with RMP ($r > 0.8$)
2. Repeatability of image analysis: we hypothesised that both inter- and intra-assessor repeatability would be similar to RMP (MDD < 10%)

5.4. Materials and Methods

5.4.1. Participants

24 participants, 17 male, aged between 4 and 15 years were recruited to the study. Participants were identified from paediatric orthopaedic clinics at the Evelina Children's Hospital. The inclusion criteria stipulated that the participants must:

1. have a diagnosis of cerebral palsy.
2. be aged between 2 and 16 years.

3. have had a 2D radiograph of the hips as part of their routine clinical management within the last two months.

The exclusion criteria stipulated that the child must not have undergone bony surgery to the acetabulum bilaterally.

Potential participants meeting the above criteria were identified by their clinical care team and their details passed to the research team. Patient information sheets and letters of invitation (Appendix 4) were sent to the families by post and a phone call at least 1 week later was made to establish whether the family were interested in the study. If interested, the participants were screened for epilepsy risk, if they were deemed ‘higher risk’ then it was necessary to arrange paediatric research nurse support at the CRF. Participants were invited to attend an ultrasound assessment within 2 months of their clinically acquired X-ray. Assessments were conducted at the CRF unless the child was already attending for a gait clinic, for these cases (N = 2), the ultrasound scan was completed in the gait laboratory at the same time as their gait assessment.

5.4.2. 3D ultrasound assessment

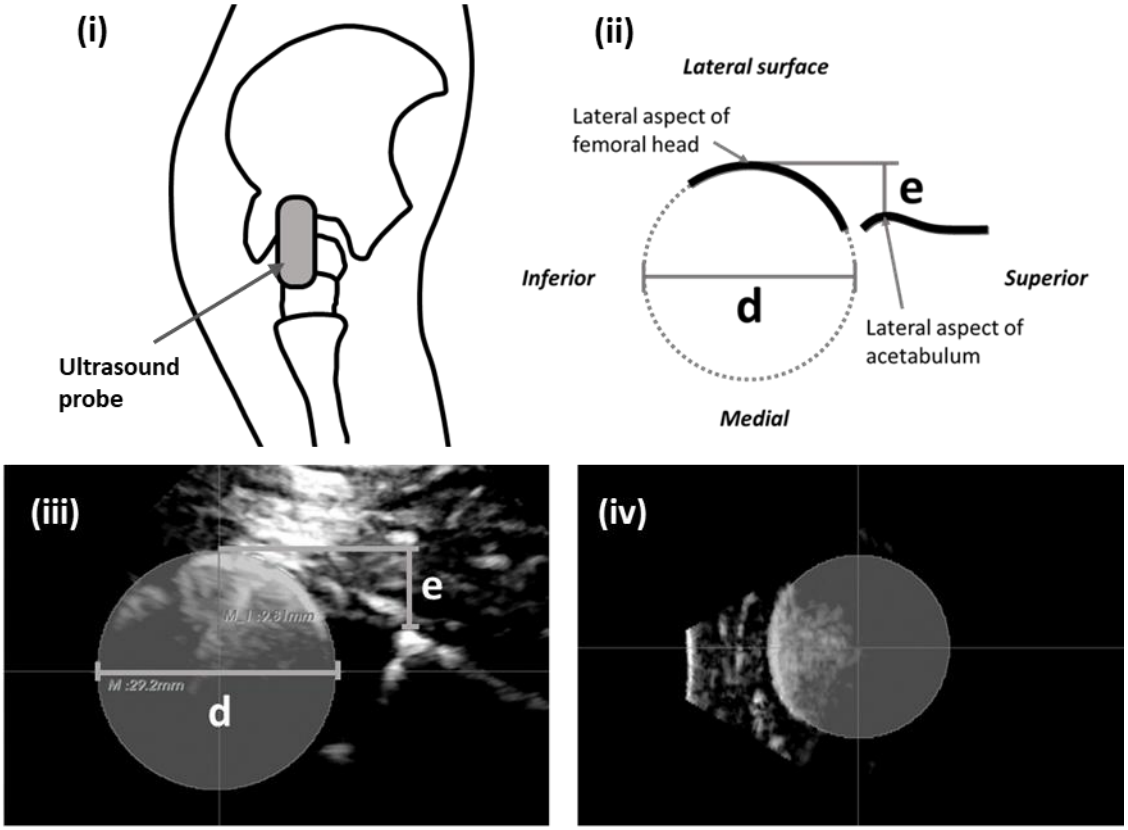


Figure 34: (i) A schematic of the probe positioning at image acquisition. (ii) A schematic of the coronal plane image showing the lateral aspect of the femoral head and the superior lateral aspect of the acetabulum. (iii) The coronal plane slice acquired inside lying showing the 'best fit' sphere. (iv) The sagittal plane slice again showing the 'best fit' sphere. a) distance between lateral aspect of acetabulum and lateral aspect of femoral head, b) estimated femoral head diameter.

Ultrasound images were acquired using either Philips EPIQ 7, with a 3D array probe (N=22), or GE Voluson with a mechanical sweep probe (N=2). The depth of the scan was set between 6 and 8cm, depending on the child's size, with a sweep angle of 60°. The child was positioned in side lying with hips extended as close to neutral as possible. The probe was orientated parallel to the superior-inferior axis of the pelvis over the lateral aspect of the hip. To optimise the image acquisition, the greater trochanter was identified, and the probe was moved posterior-superiorly to obtain an

optimal view of the femoral head and lateral acetabular border (Figure 34). Images were saved and exported in DICOM or GE .vol format.

Slicer version 4.10.1 was used for image analysis. The construction of LHC has been described previously (Chapter 4) but in summary:

1. The image volume was evaluated the volume was investigated and slices in the coronal and sagittal plane with the largest cross-sectional-area of the femoral head were chosen for analysis, see Figure 35 (iii & iv).
2. A 'best fit' sphere was fitted to the femoral head and the diameter measured as an estimate of femoral head diameter (FHD).
3. The lateral aspect of the acetabulum was identified in the medio-lateral (coronal plane) image slice, and the lateral distance between the acetabulum border and the lateral aspect of the femoral head measured.
4. The ratio of the two measurements was taken and deducted from 1 to give an estimate of the proportion of the femoral head that was covered by acetabulum, referred to as lateral head coverage (LHC), Equation 11.

$$LHC = \left(1 - \left(\frac{e}{d}\right)\right) x\ 100$$

Equation 11: Lateral head coverage (LHC) an index for quantifying the femoral head coverage in coronal plane where 'd' is the diameter of the best fit sphere and 'e' is the distance in the lateral plane between the lateral aspect of the acetabulum and the lateral aspect of the acetabulum and the lateral aspect of the femoral head.

Due to the physics of ultrasound imaging it is not possible to identify the same bony landmarks as X-ray, and thus it is not possible to create a mathematically- equivalent index. For this reason, LHC measures coverage of the femoral head by the acetabulum, and not the proportion of the femoral head that is exposed past the lateral border of the acetabulum, as is done in RMP.

It was necessary to exclude some hips from the validation study if the necessary bony landmarks could not be identified in either the X-ray or the 3D ultrasound images. Table 14 shows all participants' hips with details of inclusion/exclusion in this study. 28 hips, which met the image inclusion criteria were analysed.

Radiographs were acquired as part of the routine clinical care of the child under the standard positioning protocols used for hip surveillance. Participants were supine ensuring the hips were as close to neutral in rotation, ab/adduction and flexion/extension as the child's hip ranges allowed. The radiographers were unaware of the child's participation in a research study. RMP was measured by assessor 1, the classical method was used, i.e. using the lateral aspect of the acetabulum, due to its superior reliability[113]. All measurements were made using the image analysis package PACS Sectra IDS 7 (Version 21.1.5 2096). All RMP measures were made with at least a week's interval to the corresponding ultrasound measurements.

5.4.3. Reliability of image analysis

11 hips were selected, at random, from the first 24 hips included in the study, to investigate the reliability of the ultrasound image analysis. Three assessors with varying experience in analysing 2D B-mode ultrasound (2 months – 7 years) analysed the images. Assessor 1 (RK - myself) had six months of experience of analysing 3D ultrasound images of the hip, the other two assessors (JN and LJ) had no prior experience in analysing these images. The two inexperienced assessors underwent an hour long initial training session led by assessor 1 and had an opportunity to practice, compare and receive feedback on a training set of images prior to beginning the study. All study images were different to the training images. Each assessor used

Slicer version 4.10.1, to analyse the images. All identifying information was removed. Images were analysed across 2 sessions spaced by at least one week. In the first session each image was analysed twice by each assessor, with the image order randomised. In the second session, at least a week later, each image was analysed once by each assessor. All scores were sent to the study coordinator for compilation. One image was removed from the study as the acetabular border was consistently not visible in the slice chosen for analysis by the assessors.

5.5. Data analysis

To investigate whether the two indices were significantly different, a paired t-test was used. RMP and LHC should be inversely proportional to each other as one describes the proportion of the femoral head that is not covered by the acetabulum (RMP), and the other describes the proportion of the femoral head that is covered by the acetabulum (LHC). In order to conduct the paired t-test, 1-LHC was calculated. Pearson's correlation coefficient was used to assess the strength of correlation between RMP and LHC. SPSS (IBM SPSS Statistics for Windows, version 26) was used to perform the statistical analysis.

The inter and intra assessor repeatability of LHC measurements were investigated using intra class correlation coefficients (ICC(3,1)), SPSS (IBM SPSS Statistics for Windows, version 26) was used to compute the intra-assessor ICC (3,1) using 90 measurements (3 assessors x 10 images x 3 repeats). Inter-assessor ICC was calculated using the first of the repeat images from each assessor (a total of 30 measurements, 10 per assessor).

To investigate potential bias between sessions, the mean of the two measurements from the first session for each assessor were deducted from the second session measurements. The mean and standard error of the differences were calculated to allow calculation the upper and lower bias limits.

To aid in establishing the potential clinical utility of LHC, the minimal detectable difference (MDD) was calculated. MDD is the smallest change in two measurements that can confidently (95% confidence intervals) be taken as a true difference.

5.6. Results

5.6.1. Exclusion

Table 14 lists all recruits and details the rational for exclusion from the validation study. Out of the 48 potential hips 28 were included in the study. The most frequent reason for exclusion was that the ultrasound probe was not correctly orientated at the point of image acquisition. An amendment to the image acquisition protocol was made after recruit 9, to ensure that the same view was reliably acquired and recorded. After this amendment only 2 hips were excluded due to an ultrasound acquisition issue, 2 due to poor X-ray contrast making the acetabulum border undefinable and 2 due to the child not tolerating the test.

Recruit	*Gender*	*Left hip*	*Right hip*
1	*M*	*No orientation data recorded*	*No orientation data recorded*
2	*M*	*No orientation data recorded*	*No orientation data recorded*
3	*M*	**Included**	**Included**
4	*F*	*No imaging with probe in correct orientation acquired*	*No imaging with probe in correct orientation acquired*
5	*M*	*No acetabulum visualised*	**Included**
6	*M*	**Included**	*No acetabulum visualised*

7	*M*	*No imaging with probe in correct orientation acquired*	*No imaging with probe in correct orientation acquired*
8	*F*	*No imaging with probe in correct orientation acquired*	*No imaging with probe in correct orientation acquired*
9	*M*	*No imaging with probe in correct orientation acquired*	*No imaging with probe in correct orientation acquired*
10	*F*	**Included**	**Included**
11	*M*	**Included**	**Included**
12	*M*	**Included**	**Included**
13	*M*	*X-ray not interpretable*	*X-ray not interpretable*
14	*F*	**Included**	**Included**
15	*F*	**Included**	**Included**
16	*F*	*Child did not tolerate*	*Child did not tolerate*
17	*F*	**Included**	**Included**
18	*M*	**Included**	*Acetabulum not visualised in slice chosen as max cross-section of femoral head*
19	*M*	**Included**	**Included**
20	*M*	**Included**	**Included**
21	*M*	**Included**	**Included**
22	*M*	**Included**	*Acetabulum not visualised in slice chosen as max cross-section of femoral head*
23	*M*	**Included**	**Included**
24	*M*	**Included**	**Included**

Table 14: Table of inclusion/exclusions for each recruit and each hip within the study

5.6.2. Validation study

RMP was not significantly different to 1- LHC; t(27) = -951 (p=0.350). Figure 35 shows the relationship between the RMP and LHC. There is a strong correlation between RMP and LHC, with a Pearson's correlation coefficient of -0.792 (p<0.0001).

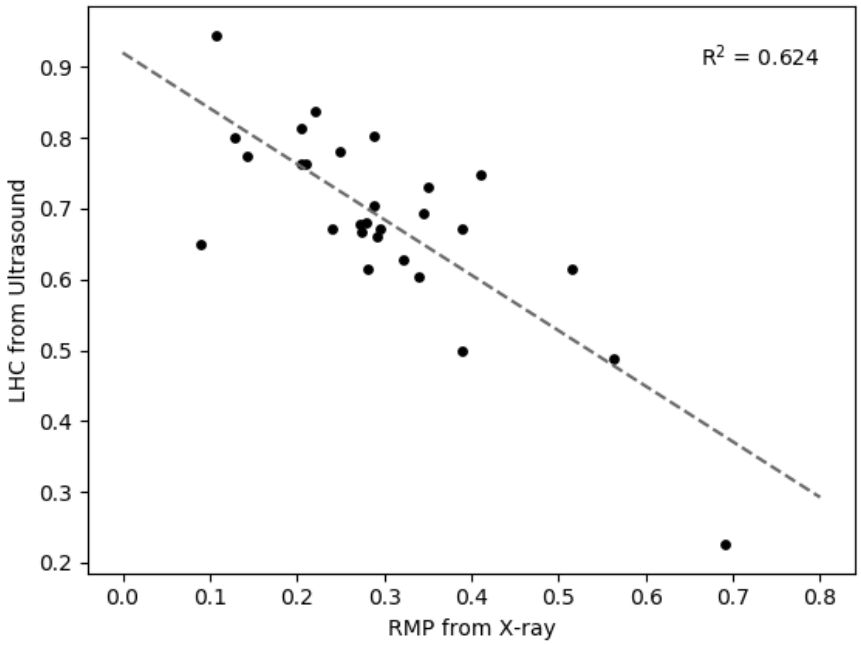


Figure 35: A comparison of RMP, measured from X-ray to LHC measurement measured by assessor 1 (RK) for 28 hips.

5.6.3. Reliability

No bias between sessions was detected for any of the assessors. The standard deviations of the averaged measurements between the assessors ranged from 0.47% for image number 3, with best agreement, to 6.99%, the worst agreement, for image number 2 (Figure 36). Inter-class reliability (ICC(3,1)) was 0.973 with 95% confidence intervals of 0.925 – 0.998 and corresponding SEM of 3.6% and MDD 10%. Intra-class reliability was 0.982 with confidence intervals of 0.967-0.991.

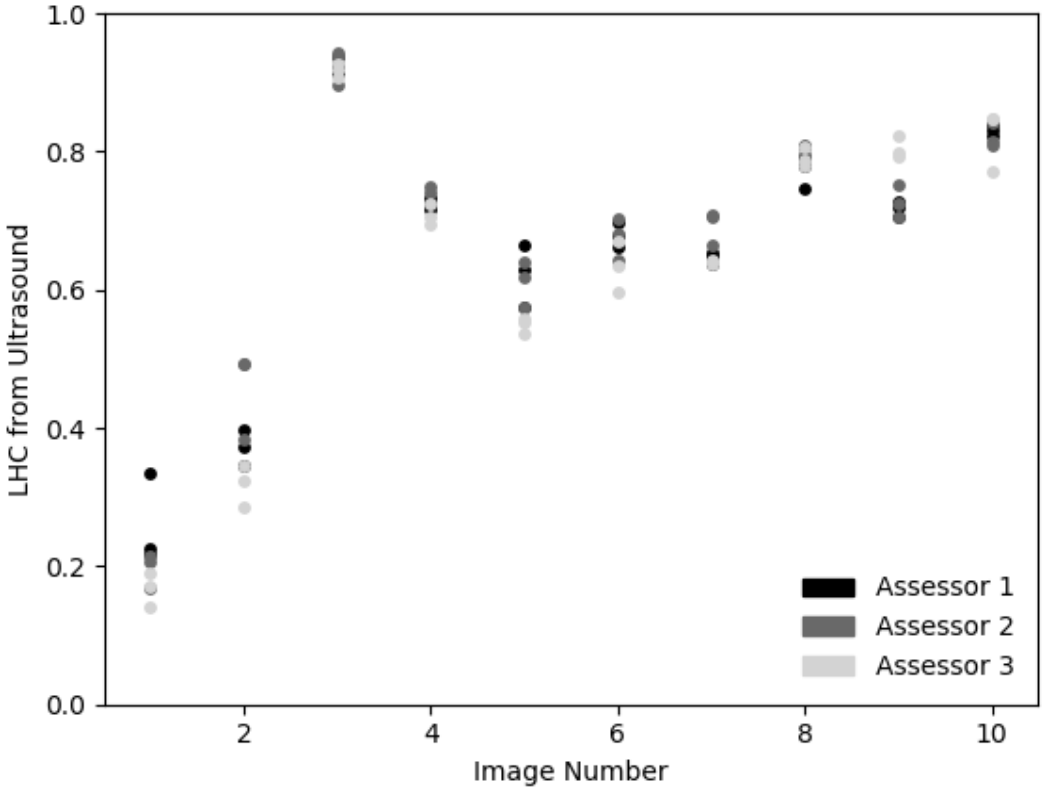


Figure 36: LHC measured 3 times for 3 assessors for each of the 10 image volumes. Assessor 3 was the most reliable with an SEM of 2.39%. Assessor 1 and 2 were very similar with SEM's of 2.91% and 2.97% respectively

5.7. Discussion

In this chapter I describe a preliminary validation of the new index, LHC, for quantifying the lateral coverage of the femoral head by the acetabulum. LHC demonstrated a strong correlation to RMP, with 62% of the variation in LHC being explained by RMP. Further, both the inter and intra assessor reliability of LHC were excellent and similar to that reported for RMP[76–78]. The SEMs and corresponding MDDs for LHC are also comparable to those reported by others for RMP (SEMs range 2.98% - 3.9%, MDDs = 8.3% - 11.5%[76,78]).

LHC is a simply constructed measure using both the sagittal and coronal planes of the ultrasound volume. Ensuring that the maximal cross-sectional area is found in two

orthogonal slices removes a source of error to which 2D radiographs are prone. LHC gives an indicator of the percentage of lateral coverage of the femoral head by the acetabulum, RMP indicates the lateral uncovering of the femoral head from the acetabulum, resulting in the inversely proportional indices. The indices are constructed from different imaging modalities that visualise the underlying anatomy in different ways, using different bony landmarks in the measurement.

This study is the first to use 3D ultrasound to assess hip development in children with cerebral palsy. There have been only a small number of studies that have used 2D ultrasound imaging in this population. Smigovec *et al*[90] visualised the hip in children with severe cerebral palsy (GMFCS IV, V) using 2D B-mode ultrasound. The scanning technique described here is an adaptation of the method used by Smigovec *et al*[90]. Smigovec *et al*[90] reported encouraging results, discriminating between measurements above and below a threshold RMP with greater than 90% sensitivity and specificity. Prior to their work, Tegnander and Terjesen[91,92] investigated the feasibility and reliability of using ultrasound to assess the fully ossified hip in children above 2 years of age. Initially they looked at 'normal hips' i.e. children with no previous hip pathology and concluded that the required bony landmarks could be visualised to measure the coverage of the femoral head by the acetabulum. They proposed normal limits for coverage (by their index) depending on age. Patients with less coverage, as measured by ultrasound, were sent for radiographs. In both these studies, the RMP result is taken to be the gold standard result and therefore confounded by errors associated with measuring RMP. For this reason sensitivity analysis, about a specific threshold, combined with a relatively small sample size has minimal validity for the new LHC index.

Previous efforts to investigate the use of ultrasound in visualising hip development in older children and specifically children with cerebral palsy may have been stalled by limitations related to inter-operator variance[93]. In contrast, 3D ultrasound is proving to be an accurate and reliable tool in the morphological evaluation of the musculoskeletal system[86], specifically in soft tissue imaging[84–86], but there are few studies of proximal femoral or hip geometry, particularly in older children. Passmore *et al*[87] used freehand 3D ultrasound to measure femoral neck anteversion angle comparing results to those obtained from MRI. The correlation was very high (Pearson correlation coefficient was 0.94) with an average difference of 1.8° between the imaging modalities across the 10 subjects. The 3D ultrasound was found to have repeatability coefficient of 3.7° with was comparable to that of MRI, which was reported as 3.1°. Geng et al[102] recently investigated the inter-rater reliability of 3D and 2D ultrasound for detecting DDH in infants under 6 months old. They concluded that 3D ultrasound had greater inter-rater reliability than 2D ultrasound and the assessments using 3D ultrasound took less time overall.

5.7.1. Limitations

Unfortunately, due to restrictions associated with COVID-19, recruitment was prematurely halted for this study and hence only 25 individuals were included. Given the preliminary nature of the study and the strong correlations demonstrated the results were still considered relevant and of interest to the clinical community. This initial work was published as an original article in Developmental Medicine and Child Neurology (DMCN) (Appendix 5).

There were a significant number of exclusions from this study, largely due to inexperience in the ultrasound image acquisition of the hip at the start of this study. Those early images not acquired with the probe orientated to be in line with the superior-inferior axis of the pelvis were not included.

5.7.2. Estimation of FHD from X-ray and ultrasound

Recalling the equations from RMP and LHC, both rely on the identification of the lateral border of the acetabulum and a measure of FHD/width. However, the two imaging modalities construct images in different ways. Planar radiographs are projection images showing areas of high and low absorption of the X-rays as they pass through the object from source to receiver, which allows for high contrast between bone (highly absorbent) and surrounding soft tissues (less absorbent) resulting normally in clear 2D imaging of hip morphology. The most lateral and most medial points of the femoral head are tracked up and the distance measured at the level of Hilgenreiner's line. This has the potential to alter the measurement depending on the anatomical position of the femoral head relative to the acetabulum. Figure 37 illustrates the potential impact of ab/adduction on the width of femoral head measurement, but rotations also have a similar impact. Ultrasound images are constructed from the reflections of the soundwaves at borders between different tissues. Bony surfaces are highly reflective to these soundwaves and as such ultrasound cannot visualise structures that sit deeper to a bony surface.

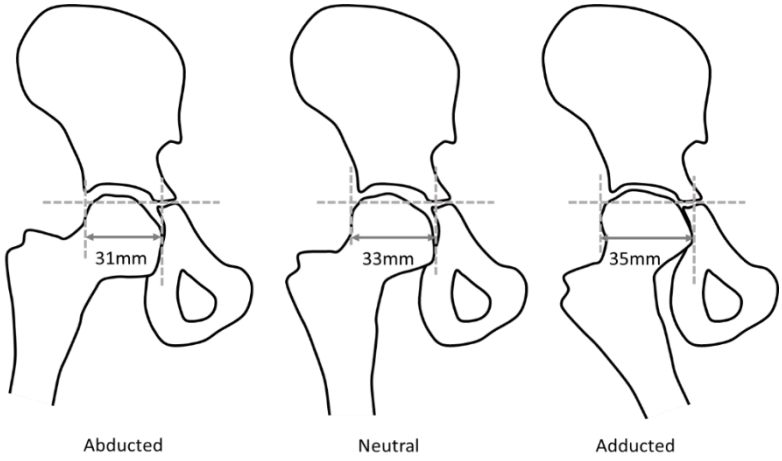


Figure 37: Schematic showing the impact of ab/adduction on estimate of FHD in RMP measurement

In ultrasound, both the estimation of FHD, and to a lesser extent the slice selected for analysis, are dependent on the lateral curvature of the femoral head. The estimate of FHD was made by fitting a sphere to the curvature of the lateral aspect of the femoral head at maximal cross-sectional area in both the sagittal and coronal plane, by eye, using image analysis software Slicer. Other methods were investigated and described later in the chapter.

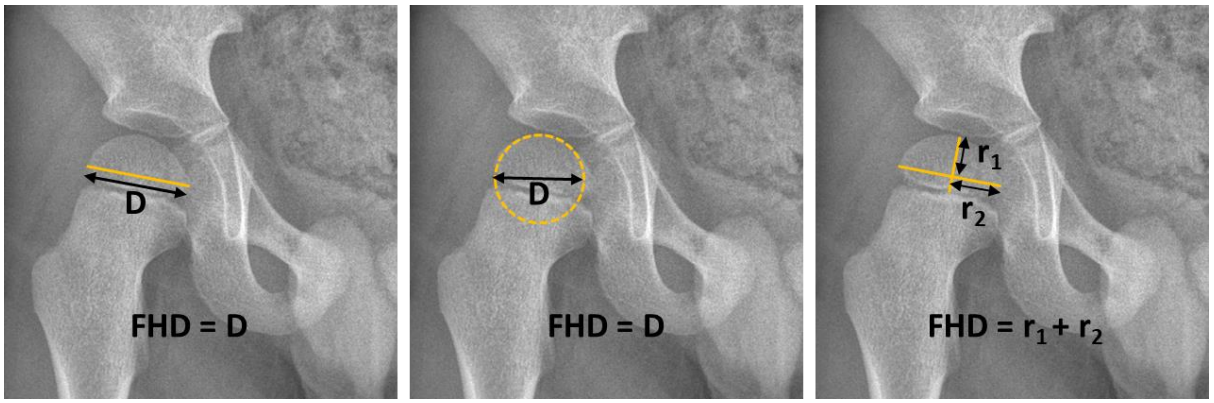


Figure 38: Estimates of FHD from A-P planar X-ray. (i) The maximal diameter. (ii) 'Best fit' sphere fitting to the lateral surface. (iii) The averaging method, the estimate is made from 2 perpendicular radii.

The femoral head is not truly spherical and thus both the slice selection and the size of the sphere are subject to user interpretation of ‘best fit’, causing potential

discrepancies. To increase confidence in the FHD estimates from ultrasound, they were compared to those derived from X-ray. Three different estimates of FHD from X-ray were conducted, the first, a simple measure of the widest part of the femoral head from the anterior-posterior X-ray (Figure 38(i)). The second, the diameter of a circle fitted to the lateral curvature of the femoral head, similar methodologically to the estimate of FHD from ultrasound (Figure 38(ii)). Thirdly, an average of two orthogonal diameters. To construct this, the maximal diameter is found and then a perpendicular bisector drawn. The sum of the two perpendicular radii is the estimate of FHD (Figure 38(iii)). We believe that this third method will be the most robust measure to estimate the average diameter for the femoral head given the non-spherical nature of the head.

No systematic difference between the ultrasound and this estimate of FHD was detected (Bland Altman Figure 39). Table 15 compares the absolute difference between the FHD estimated from ultrasound and the different estimates from X-ray as well as the error between the different X-ray techniques. The mean absolute % error between X-ray and ultrasound measurements we computed to be around 10% (Equation 12). It should be recognised that this error may be made up from multiple sources including the variance owing to the error associated with measurements made on both the X-ray and ultrasound images.

$$mean\ absolute\ difference\ (\%) = 100 * \left(\frac{\left| FHD_{ultrasound} - FHD_{x-ray} \right|}{(FHD_{x-ray})} \right)$$

Equation 12: Equation for calculating the absolute percentage error between two methods.

	Comparison of Ultrasound and X-ray FHD			*Comparison of different X-ray FHD measures*		
Average difference (%)	Ultrasound vs max point	Ultrasound vs sphere fitting	Ultrasound vs 2 radii	Max point vs sphere fitting	Max point vs 2 radii	Sphere fitting vs 2 radii
	10.74%	11.90%	10.49%	13.71%	7.02%	7.08%

Table 15: A table comparing the average absolute differences between different estimates of FHD across 24 hips between the 4 different estimates (3 X-ray methods, 1 ultrasound method)

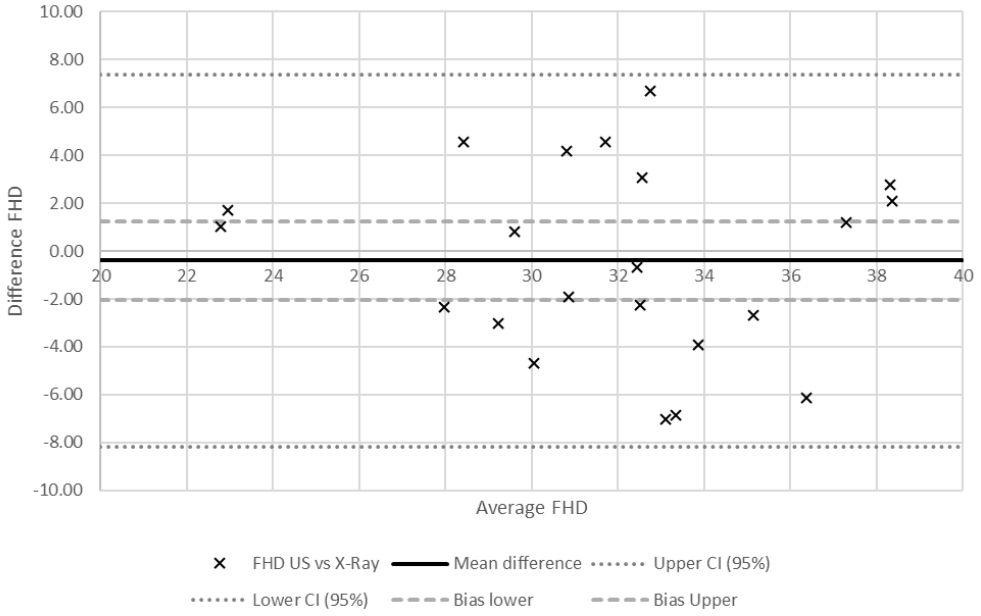


Figure 39: A Bland-Altman plot comparing the ultrasound estimates of FHD to estimates from X-ray (2 radii method) for 24 hips

5.7.3. Alternative ultrasound FHD estimates

There are many complex and computationally intense methods of best fitting a sphere to an arc. Each is a compromise between complexity, data distribution and robustness to outliers. For two main reasons it was felt that complex techniques were not required in this circumstance. Namely the femoral head is not a sphere and therefore the improvement in accuracy of these computational methods may not be substantial.

Secondly, this technique was developed with the aim of possible adoption into clinical service, if there is a requirement to run mathematical transformations or complex functions on the data to generate a result it will provide a barrier to adoption.

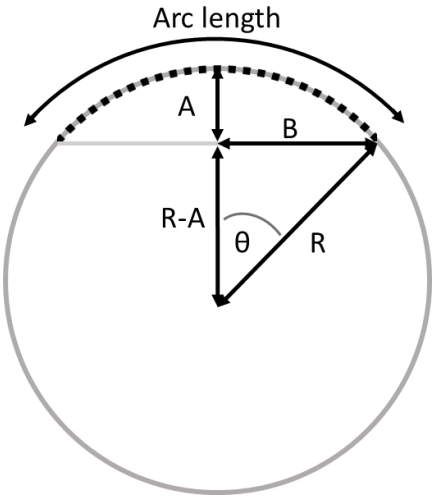


Figure 40: Schematic of relationships between arc length, chord and radius

A simple alternative method was derived, using the mathematical relationships between the radius, arcs and chords. Once the slice of maximal area was found a chord was drawn on the coronal plane slice. The perpendicular bisector to the chord was then drawn (Figure 41). From these two measurements it is possible to compute the radius of the circle that the chord is drawn through (Equation 13).

$$R = \frac{B^2 + A^2}{2A}$$

Equation 13: Defining the relationship between the radius (R) and the chord length (2B)

The FHD was estimated on 24 hips, the cohort that were included in the research paper, and the corresponding LHC calculated. The correlation between LHC calculated with different FHD and RMP we computed.

Figure 41 shows that the LHC calculated using the FHD via the chord method showed a marginally stronger correlation with RMP in comparison to the 'best fit' sphere method. However, the absolute difference between the FHD estimates from the ultrasound chord method compared to the X-ray estimates were greater (Table 16).

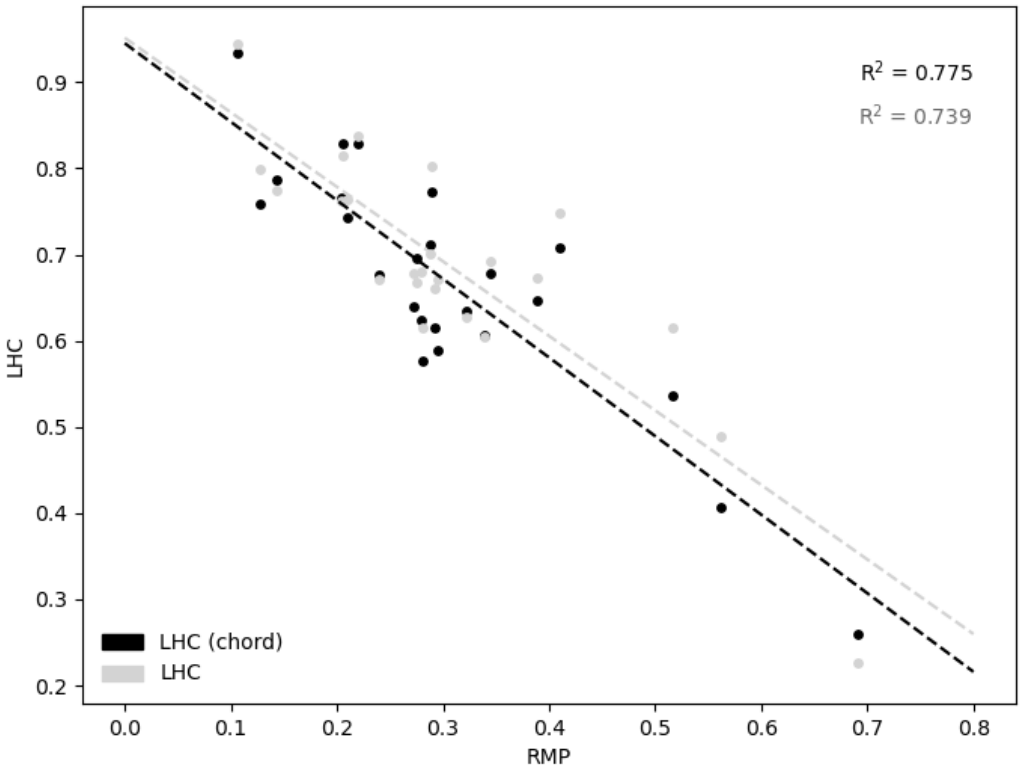


Figure 41: Comparison between LHC using the two different methods for estimating FHD

	US (sphere) vs US (chord)	***US (sphere) vs X-ray (2 radii)***	***US (chord) vs X-ray (2 radii)***
Average absolute % difference	9.1%	10.1%	14.4%

Table 16: Results from the comparisons between the different FHD estimates

The 'best fit' sphere, remains likely to be the simplest method of generating an estimate of FHD, given the sphere is needed to find the slice of maximal area in the two orthogonal planes. However, the proportion of the lateral surface of the femoral

head visible, will affect to confidence with which a ‘best fit’ sphere can be fitted. By implementing the chord method, an investigation into the effect of the proportion of the arc length visible in the chosen slice could be conducted. The arc length, as a proportion of the total circumference of the sphere was computed Equation 14.

$$\%\, FH\ visible = \frac{Arc\ length}{\pi d} \times 100$$

Equation 14: Defining the proportion of the femoral head visible in the Ultrasound image, where 100% would indicate the full circumference of the femoral head could be visualised.

Figure 42 shows the distribution the RMP vs LHC categorised by the proportion of the femoral head that was visible. There is not enough data to conduct a full analysis into the effects of the visible arc length but it was observed that the more displaced hips tended to have a greater proportion of the arc visible.

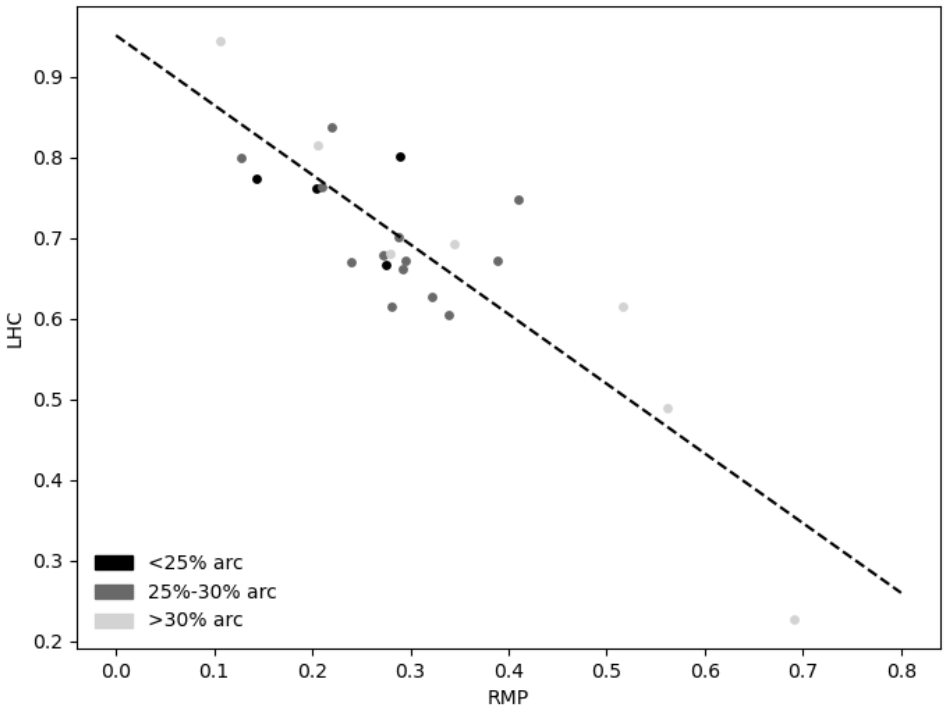


Figure 42: Investigating the effect of the proportion of femoral head visible and correlation of LHC and RMP measurements

Whilst, the estimate of FHD from 3D ultrasound has limitations, particularly in those with significant hip dysplasia where the femoral head can be significantly deformed, the challenges are outweighed by the advantages of facilitating an index that is constructed as a ratio. Mainly for the purposes of normalisation, but also to prevent any potential issues with scaling in the image. Scaling problems can arise as technical issues, for example if an ultrasound machine has not been adequately calibrated or there is an incompatibility between the raw data files and the chosen image viewers for interpretation. Normalisation is essential to accommodate structures of different sizes, for example a femoral head with a lateral surface protruding 10 mm past the acetabulum is well covered if the femoral head is 40 mm, but it is cause for concern if the femoral head is only 25 mm in diameter.

5.7.4. Challenges at image acquisition

At image acquisition, the greater trochanter is identified, the probe is then moved superiorly and posteriorly and centred over the exposed portion of the femoral head. This technique was based on the technique reported by Smigovec *et al*[90] but adapted for ease of application. The greater trochanter can be palpated easily and therefore reliably identified. LHC, is therefore, more of a measure of posterior-lateral coverage as the image is not acquired from a pure lateral position. Further, the position of the greater trochanter varies between individuals. In the typically developing paediatric population the average anteversion is age dependent ranging from 32° at the age of one to 16° at the age of sixteen, the standard deviation within an age range is approximately 7°[114]. However, in children with cerebral palsy the variation is

significantly greater with GMFCS levels III to V having femoral anteversion angles of approximately 40° on average (range 25° to 67.5°)[115].

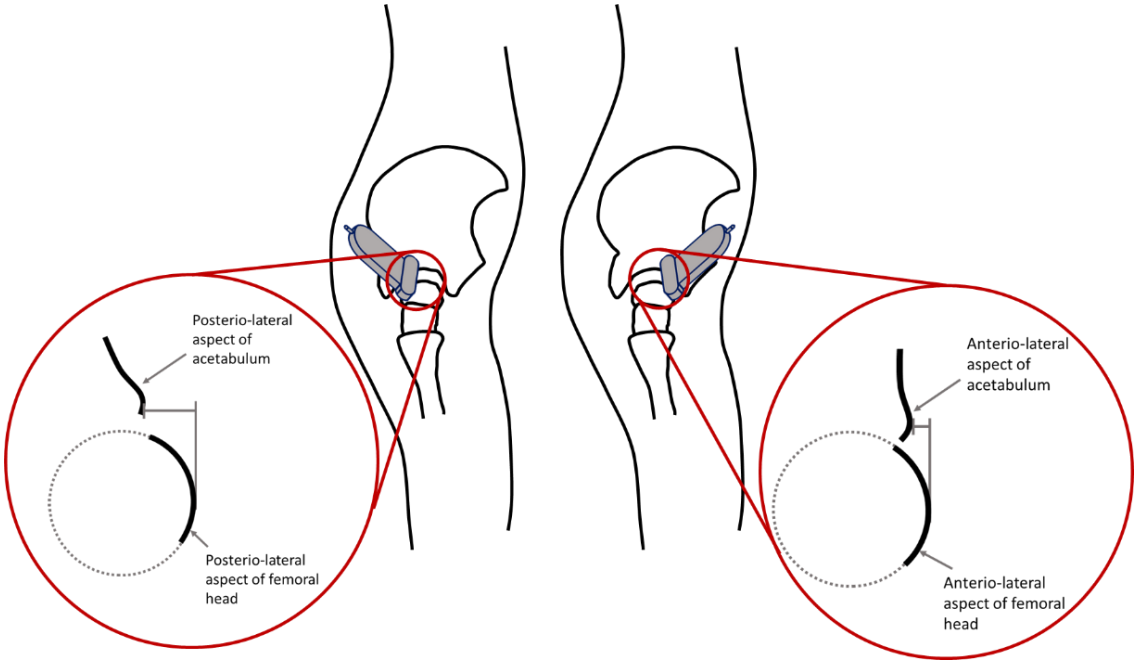


Figure 43: Schematic showing extreme probe positions and the effect on the identification of acetabulum border

The errors at acquisition of the 3D ultrasound volumes have not been investigated thoroughly. The 3D nature of the ultrasound moderates errors associated with probe positioning and orientation associated with 2D ultrasound[102], however using a variably positioned bony landmark to identify the probe position will result in variability in the ultrasound volume acquired. The plane of the chosen slice of maximal cross-sectional area is dependent on the probe positioning. The identification of the acetabular border is also dependent on slice selected as it is defined from the chosen coronal plane slice. Figure 43 illustrates, at the extremes, how this variability could affect image acquisition.

Without a better knowledge of the bony development of each of the recruits' hips, it is hard to account for developmental or other factors that might influence LHC, but they should be considered as a potential source of error. In this study probe position and femoral anteversion were not recorded and therefore their influence cannot be quantified. In individuals with high levels of femoral anteversion LHC may be underestimated, particularly in cases where the posterior surface of the acetabulum is under-developed (see Figure 43) [46,116] . The plane of the image is further defined by the positioning of the probe perpendicular to the skin surface. In the sample population, this was considered acceptable as there was minimal excess tissue coverage and 3D tracking of the anatomy and the probe (say, using motion capture technology) was deemed un-realistic in the clinical environment. Care was taken to ensure that the probe was held perpendicular to the skin surface and consistently just posterior to greater trochanter.

5.7.5. Sensitivity of LHC to out of plane displacements

The *in vitro* experimentation exposed the sensitivity of LHC to out of plane displacements. The impact of this sensitivity is unquantified in this study as there was no control 3D imaging available. This may explain some of the discrepancy between LHC and RMP. We know from the work of Brunner *et al*[25] that in approximately half of individuals there is a significant out-of-plane displacement. Whilst LHC's sensitivity to these non-lateral displacements would weaken the correlation between RMP and LHC, it means that it would be capable of identifying displacement in individuals that might otherwise not have been detected by RMP measures from planar X-ray. This is explored further in Chapter 6.

5.7.6. LHC and the characterisation of hip dysplasia

Hip dysplasia is a mal-development of the hip in three dimensions and as such would ideally be monitored by 3D imaging. Traditional 3D imaging modalities are not practical options for routine monitoring of hip development in children with cerebral palsy.

Although LHC is in effect a 2D measurement made from a single ultrasound image, 3D imaging allows the user to select the optimal image from the acquired volume.

For further validation, a comparison between LHC and a similar measurement derived from 3D imaging, where error sources could be minimised and the comparative measurements could be considered a true gold standard, is needed. Gose *et al.*[46] compared a CT derived index to RMP and reported a strong correlation (r = 0.85, P<0.0001) (figure 5[46]) between the measurements. Their results were comparable to the agreement found in our study (r=0.79) between LHC (from ultrasound) and RMP (from X-ray). It would be of interest to compare 3D ultrasound to another 3D imaging modality, such as MRI, CT or new bi-planar X-ray imaging modality, EOS®. Neirynck *et al*[83] have shown that RMP measured from standing radiographs using EOS was statistically similar to RMP from standard supine planar X-ray. This technology has the capability to thoroughly assess and quantify the projection error associated with RMP. Investigation of the 3D reconstruction from EOS re-sliced to derive the 2D planar X-ray at different rotations, would provide insight into the magnitude of the errors at image acquisition for RMP. Better understanding of the sensitivity of RMP to projection angle would allow for quantification of the impact of the RMP variation on any future comparisons to alternative imaging modalities.

5.8. Clinical implications

The use of ultrasound to evaluate the hip in young infants has transformed the screening of developmental hip dysplasia (DDH)[117]. Like hip dysplasia in CP, the definition of DDH is not completely agreed upon, however it is widely accepted that DDH refers to “a continuum of abnormalities in the immature hip that can range from mild dysplasia to dislocation”[118], in otherwise healthy infants. Ultrasound lends itself well to imaging of the hip in the very young as the hip has not ossified and therefore sound waves are able to partially penetrate through the hip joint allowing visualisation of the acetabulum. As the hip ossifies, it becomes impossible to get the same clear images of the joint. However, as this study has shown it is still possible to visualise significant anatomical landmarks and make measurements of hip geometry which may have diagnostic value.

Using 3D ultrasound imaging would allow for more frequent and repeated assessments to be performed because ultrasound is a non-ionising imaging modality. Ultrasound imaging would also allow for the hip to be imaged in different positions, providing further information about the hip development that is not currently collected from single radiographs. Repeated measurements in the same position would also allow for greater confidence in a measurement. Further, screening programmes often do not have frequent monitoring for the less affected children with cerebral palsy as they are less at risk of hip displacement. Depending on the programme individuals may be discharged after a single ‘normal’ radiograph or receive a further X-ray at around the age of 8 years (after which very few hips go onto dislocate[28,30,32]). Kentish *et al*[97] reviewed 1115 children who had been engaged in their hip screening programme. Of these, 28% had RMP of greater than 30%. In this group with high

RMPs, 16% were GMFCS level I or II. Using a non-ionising imaging modality such as 3D ultrasound would allow for safe continued monitoring in the more able group.

5.9. Conclusion

In this chapter I present an initial concurrent validation of the use of 3D ultrasound in monitoring hip development in children with cerebral palsy. The results show that LHC is comparable to RMP in estimating hip dysplasia with similar levels of inter and intra assessor reliability to those reported for RMP. With the potential to increase assessment frequency the 3D ultrasound assessment technique could, as a minimum, provide a non-ionising alternative for monitoring hip dysplasia in cerebral palsy. It is also likely that the additional structures and views that can be imaged with ultrasound compared to a 2D radiograph could provide valuable information on hip management for individuals with cerebral palsy (Chapter 6). Further investigations are required to appreciate the full potential of 3D ultrasound in the monitoring of hip dysplasia in children with cerebral palsy.

6. The measurement in vivo of the posterior displacement of the femoral head in children with cerebral palsy using 3D ultrasound.

6.1. Overview

In this chapter I present a further investigation of the use of 3D ultrasound for evaluation of hip development in an *in vivo* setting. The aim of this work was to investigate the potential value of the measurement of displacement in the sagittal plane in the assessment of hip dysplasia.

Here, we evaluate *in* vivo an index that we investigated in the *in vitro* work described in Chapter 4. Femoral head posterior position ratio (FHPPR) was measured on anterior scans acquired as part of the clinical study. The image volumes were acquired with the participants lying in supine, and the ultrasound probe positioned over the anterior surface of their hips

From other studies[25,46], we know that hip dysplasia is not a simple lateral displacement of the femoral head relative to the acetabulum. From our *in vitro* experimentation we understand that LHC, derived from 3D ultrasound assessment is sensitive to displacements in the sagittal plane. We suspect, based on results published in the literature, some antero-posterior displacement would have been present in a significant proportion of our study participants. FHPPR was used to quantify this component. The results were compared to others reported in the literature, which used different imaging modalities and clinical measurements. No X-ray imaging in this plane

was acquired and therefore, it was not possible to directly compare FHPPR with an X-ray derived index.

In vitro, FHPPR was insensitive to medio-lateral displacement of the femoral head, so we combined image modalities to explore hip displacement trajectories in the transverse plane. RMP (from a coronal plane radiograph) and FHPPR (from a 3D ultrasound volume) were used to plot the hip displacement trajectories for each of our recruits. The results showed that in approximately 42% of our participants, there was a deviation from a pure lateral displacement of at least ±13°.

These measurements indicate that a significant component of displacement in current clinical imaging assessment is unaccounted for. FHPPR, from 3D ultrasound volumes, in combination with coronal-plane X-ray imaging, may improve identification of individuals at risk and selection of appropriate interventions. Further investigation is warranted.

6.2. Introduction

The aetiology of hip dysplasia in children with cerebral palsy has been widely investigated, both in clinical studies and through mathematical modelling. Probably the most influential clinical study was conducted by Reimers in 1980[27], leading to the development of Reimers migration index, (RMP). He evaluated different types of hip displacement and concluded that the primary causative mechanism was the abnormal forces generated by the hip adductor muscles, with the hamstrings and iliopsoas muscles having a secondary influence. The widespread use of RMP in the clinical setting has led to investigations and surveillance programmes that quantify lateral displacement of the hip.

However, by investigating the morphology of the roof of the acetabulum in children with CP, undergoing hip reconstruction, using CT scanning, Brunner *et al*[25] observed clear channels in the roof of the acetabulums through which the femoral heads had displaced. By analysing the direction of these channels, they observed huge heterogeneity in the direction of displacement across the cohort. The median direction was 2° posterior to purely lateral or purely coronal plane displacement. However, the direction of hip displacement varied from 33 degrees anterior to 70 degrees posterior of a purely lateral direction with over 50% of the 24 hips studied having a deviation from a purely lateral displacement of greater than 13 degrees. Gose et al[46] conducted a CT study looking at the positioning of the femoral head relative to the centre of the acetabulum. They found that in 82% of the hips the femoral head was located posteriorly, superiorly as well as laterally relative to the centre of the acetabulum.

The musculoskeletal modelling of Miller *et al*[26] suggests forces acting on the hip in the child with CP are both high in magnitude and altered in direction. They estimated that

the hip of a child with CP in a typical resting position (knee flexed (60°), hip flexed (50°), adducted (30°) and internally rotated (50°)) may have 3 times the forces of their unaffected peers. Further, they calculated that if the hip of the child with CP was placed in a neutral position in their simulation, the magnitude of the force would double again. What is more, these excessive forces were directed posteriorly, superiorly, as well as laterally, whilst the forces from the 'normal' hip model were directed medially and superiorly. The direct of the force in Miller *et al*'s model[26] is similar to that observed by Gose *et al*[46] in their CT morphometric analysis of the acetabulum in children with CP. The magnitude of the forces reported will be highly sensitive to the input parameters of the model, which included an assumption of 40% reduction in muscle length in the CP model. As such, comparing the absolute results has limited validity, but the direction and presence of the differences between the states (neutral positioning and 'spastic' positioning) is likely to be relatively robust.

The clinical implications of this study were to avoid the use of bracing to correct hip position in the child with CP, and that femoral anteversion and femoral neck shaft angle may be a result of excessive forces rather than a causative factor in hip displacement. Although modelling studies may suffer from many assumptions including the contributions of individual muscles and simplification of joint morphologies, this study does suggest that large deviant forces in the hip may encourage displacement in a direction which is not purely lateral. This modelling, alongside clinical studies[25,27,46,116], highlight the need for multi-directional imaging to assess hip dysplasia.

Hermanson et al[29] defined a predictive equation, based on measurements taken from X-ray (RMP and head-shaft angle (HSA)) and the presentation of the child (age and GMFCS level) which aimed to predict an individual's likelihood of developing severe

hip displacement within the following 5 years. The risk score runs from 0-100% likelihood in bands of 10%. The team reported an accuracy of 86% for their predictive equation, titled the CPUP score. Clinical examination data collected at the time of assessment was not included in the equation. According to the authors (personal communication) the inclusion of clinical examination data did not improve prediction.

The empirical data quantifying hip displacement progression rates are all derived from 2D radiographs. Where more 3D imaging has been used, the data is cross-sectional and therefore does not adequately inform predictions of hip dysplasia. However, 3D investigations are likely to have a significant impact on the clinical management of the individual patient in terms of informing specific surgical procedures, evaluating an individual's risk of developing further significant hip displacement and quantifying the outcomes of different preventative treatments. 3D ultrasound data could provide a modality capable of quantifying 3D hip displacement without repeated exposure to ionising radiation.

The results from the in *vitro* work reported in Chapter 4 indicated the potential of an index (FHPPR, Equation 15) for assessing the positioning of the femoral head in the anterior-posterior direction, or sagittal plane. The index showed high sensitivity to true pure posterior displacement whilst being significantly less sensitive to true lateral displacement.

6.2.1. Hypothesis

Informed by the work of Brunner[25] and by Miller[26], we hypothesised that a number of children in our study will have displacements which have significant component in the anterior or posterior direction.

6.3. Materials and Methods

Ultrasound images were acquired using either Philips EPIQ 7, with a 3D array probe (X5-1 Phased Array Probe), or GE Voluson with a sector sweep probe (GE Healthcare RM6C probe). The depth of the scan was set between 6 and 8cm, depending on the child's size, with a sweep angle of 60°. The child was positioned in supine with hips mildly flexed, rotated and ab/adducted as close to neutral as comfortable. As with the laterally-acquired images, the probe was orientated parallel to the superior-inferior axis of the pelvis, this time over the anterior aspect of the hip. To optimise the image acquisition, the pelvis was palpated to identify the hip joint, and the probe was manipulated so that the largest surface of the femoral head could be seen, ensuring that the superior acetabular border was in the frame (Figure 44). Images were saved and exported in DICOM (EPIQ-7) or GE .vol (Voluson) format.

As with the measurements of lateral head coverage (LHC) (cf. Chapter 5), Slicer version 4.10.1 was used for image analysis. The volume was investigated and slices in the coronal and sagittal plane with the largest CSA of the femoral head were chosen for analysis see Figure 45 (iii & iv). A 'best fit' sphere was fitted to the femoral head and the diameter measured as an estimate of femoral head diameter (FHD). The most posterior element of the inferior point of the anterior aspect of the acetabulum was also identified in the same sagittal plane image slice, and the distance between the acetabulum border and the centre of the estimated femoral head in the anterior-posterior plane was measured (Figure 44 (ii)).

The femoral head posterior position ratio (FHPPR) was calculated by deducting the radius of the estimated femoral head from 'a', the distance between the acetabular border and the central line of the estimated femoral head (Equation 15). A positive

value here indicates that the anterior surface of the femoral head does not protrude past the anterior acetabulum border, i.e. it is posterior to the acetabulum. A negative value indicates that the anterior aspect of the femoral head is anterior to the acetabulum border. To normalise to the size of the child's hip, this value is divided by the estimated femoral head diameter. Larger numbers, positive or negative, may indicate abnormal posterior or anterior hip positioning.

$$FHPPR = \frac{a - 0.5d}{d}$$

Equation 15: Index describing the femoral head position relative to the anterior acetabulum edge in the anterior-posterior plane

Only hips included in the study of LHC were included in this analysis. It was necessary to exclude five of anterior hip images, where the required image volumes were either not collected or of insufficient quality to define the anterior acetabulum border. A total of nineteen hips were included in this analysis.

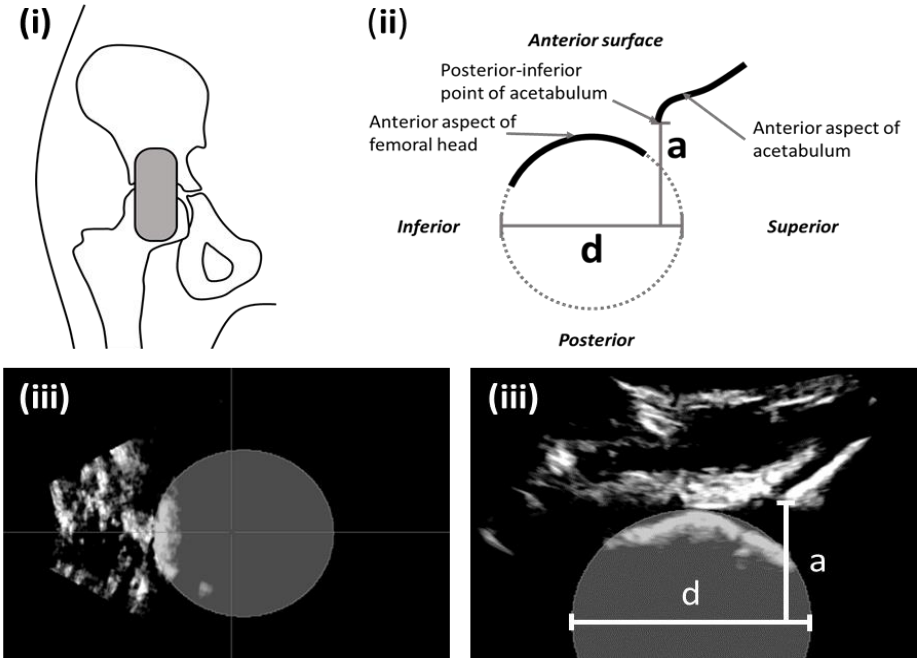


Figure 44: A schematic of the probe positioning at image acquisition. (ii) A schematic of the sagittal plane image showing the anterior aspect of the femoral head and the posterior- inferior aspect of the acetabulum. (iii) The coronal plane slice acquired in supine showing the 'best fit' sphere. (iv) The sagittal plane slice again showing the 'best fit' sphere. a) distance between posterior- inferior aspect of acetabulum and mid-line of femoral head, d) estimated femoral head diameter.

6.3.1. Interpretation of Ultrasound Images

The FHPPR index was developed in the *in vitro* setting, where identification of the landmarks is straight forward as there is no soft tissue reflecting the ultrasound. In many cases, in the *in vivo* setting, the echo from the acetabulum border is of similar intensity to that from the labrum and other surrounding soft tissues. The interpretation of the image volumes was informed by studies investigating the use of 2D ultrasound in the assessment of anterior labrum tears[119–122] and a study of 2D ultrasound in hip dysplasia in children with cerebral palsy[90], knowledge of hip anatomy, and an understanding of the construction of ultrasound images.

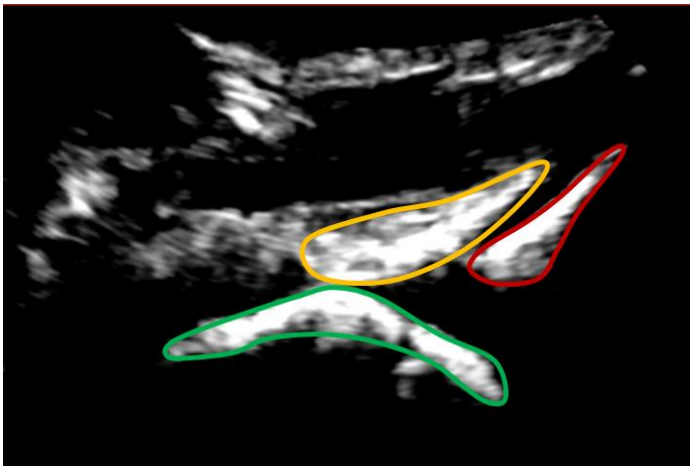


Figure 45: Identification of significant anatomy from sagittal plane ultrasound slice. Red = acetabulum, yellow = labrum, green = femoral head.

In the imaging of the anterior aspect of the hip with ultrasound, the labrum represents a highly echogenic structure which may be confused with a bony surface. In an attempt to define clearly the surface of the FH in the presence of the labrum, image contrast was increased. This is illustrated in Figure 45. Shadowing posterior to the acetabulum is observed, indicating that all the sound is nearly completely reflected at this bony surface. The labrum, which sits superficially to the femoral head, has a strong echo,

but the deeper lying femoral head can still be visualised, indicating that the labrum causes only partial refection of the incident ultrasound energy.

6.4. Data analysis

It was hypothesised that the distribution of displacement direction in our study would be similar to that reported by Brunner *et al*[25]. They displayed the distribution of displacements as a box plot, it is therefore not possible to fully analyse the distribution of displacements across their cohort. However, by taking manual measurements from the box-plot displaying their results it was possible to estimate the inter-quartile ranges and re-plot their results (Figure 46). The scale was also inverted so that negative numbers represented anterior displacement and positive numbers posterior displacement for consistency with this study. This analysis showed, over 50% of the hips studied fell outside -13° to +15° of pure lateral displacement. We hypothesised that the direction of displacement of the hips in our study would follow a similar distribution.

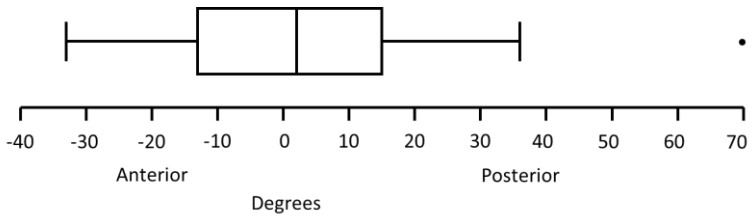


Figure 46: Recreation of Brunner's results, distribution of the direction of dislocation relative to pure lateral in degrees

To test this hypothesis a Chi squared test[123] was used to test for a difference between the proportion of hips that displaced within a +/- 13° envelope of a pure lateral direction, and those that displaced outside this envelope, between this study and that of Brunner (Table 17).

To assess lateral hip displacement, we used RMP rather than LHC due to the latter's dependence on out-of-plane displacement. The estimation of the direction of femoral head displacement is dependent on the assumption of orthogonality between RMP and FHPPR.

6.5. Results

The direction and magnitude of the hip displacement for each hip studied was displayed as a vector (Figure 47). As a guide, lines representing deviations at 10 degree increments from 30 degrees anterior to 30 degrees posterior are included (dotted lines Figure 47). Additional lines are added indicating the 50th percentile posterior and anterior range of deviation from a purely lateral displacement from the CT study of Brunner *et al*[25]*.*

From Figure 47, two clusters of vectors with greater than 13° deviation from pure lateral in both anterior and posterior directions can be identified. In each cluster there are 4 individuals. There is 1 hip that has greater than 20° deviation in the anterior direction and 3 hips that display greater than 20° deviation in the posterior direction. 42% of hips in this study had greater than 13° of either anterior or posterior deviation from lateral. 21% of hips had greater than 20° deviation from the lateral axis and 10.5%

(2 hips) had greater than 30° deviation from the lateral axis, both in the posterior direction.

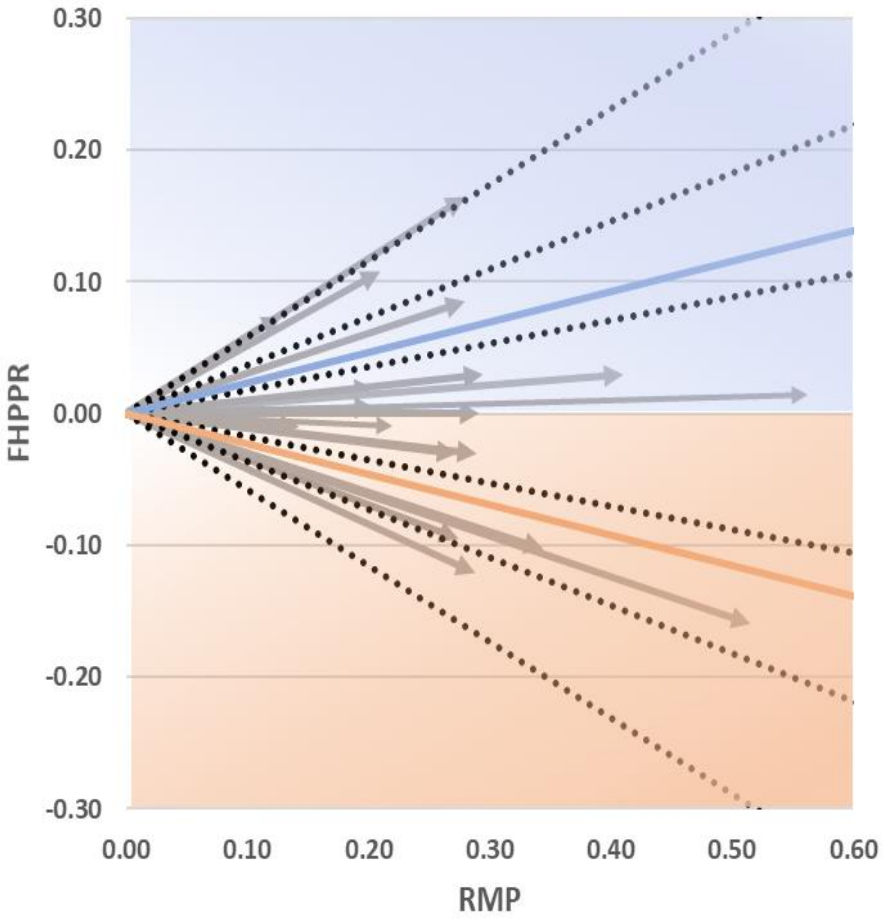


Figure 47: Hip displacement direction and magnitude for each of the individuals in the study. The vertical axis is the anterior-posterior direction, with the blue zone indicating posterior displacement and the blue line representing 13° posterior displacement. The orange zone indicates anterior displacement and the orange line represents 13° anterior displacement. The hashed lines represent +/- 10°, 20° and 30° deviation from true lateral. The horizontal axis is the medial-lateral direction. The greater the magnitude of the arrow the more displaced the greater the hip displacement. RMP (from X-ray) and FHPPR (from 3D ultrasound) are assumed to be orthogonal for this analysis.

A 2x2 contingency table was constructed with Brunner's study data and our study data (Table 17). The χ^2 was calculated and the probability distribution table[124] was then used to look-up the result, to establish the p-value and hence the significance of the result. We returned a $\chi^2 = 0.266$, which gave a p-value of greater than 0.5 (one degree of freedom), indicating that there was no significant difference between the proportions of displacement either side of the +/- 13° envelope.

When the total number of samples included in the analysis is below 100, the Yates[123] correction can be used for a better approximation of the p value. This yields χ^2 = 0.0431, and confirms that there is no significant difference between the two studies.

Study	*Within +/- 13° envelope*	*Outside +/- 13° envelope*	*Total*
Brunner et al	12	12	24
Our study	11	8	19
Total	23	20	43

Table 17: A 2x2 contingency table for Chi squared statistical test, comparing the displacement directions from Brunner et al's study to our study

6.6. Discussion

Our results support our hypothesis that there was no significant difference found between the distribution of the direction of displacement of the hips in our study and Brunner's study, when analysed using a threshold of +/-13° from pure lateral. In our study, 42% of hips were outside of the 13° either side of lateral envelope, whereas 50% of hips fell outside of this envelope in the Brunner study.

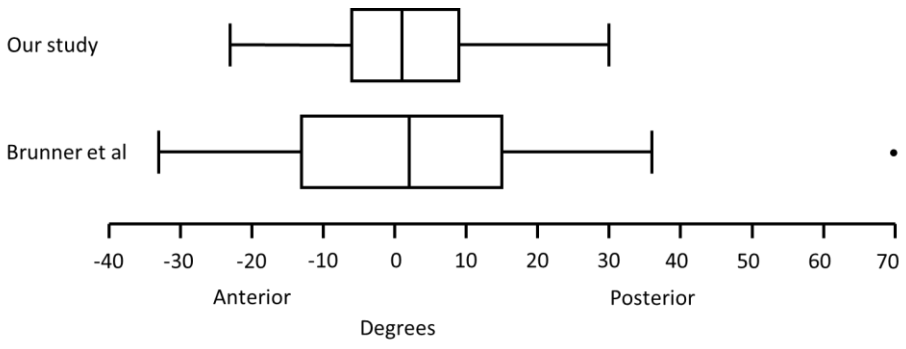


Figure 48: A comparison of our study results and the recreation of Brunner's results, distribution of the direction of dislocation relative to pure lateral in degrees

In both the study reported and from the work of Brunner, a significant number of hips (about 10%) were displaced posteriorly at a deviation to a purely lateral displacement

of greater than 30°. A deviation of 30° implies a posterior displacement of 50% of the magnitude of the total displacement. These studies highlight the need to perform multi-planar clinical studies of hip displacement. The presence of a large posterior hip displacement may have disproportionate clinical significance.

Figure 48 shows our data displayed as a box-plot against the re-created, and inverted, results from Brunner *et al*'s study[25]. Brunner's trajectories of hip displacement were more variable than the ones we recorded here using 3D ultrasound. This might be for one or more reasons. There could be systematic differences between the measurement techniques, the methodologies for establishing and measuring the direction of displacement are not comparable between studies, and therefore absolute values cannot be compared. But it may be due to the sample population in the studies with those waiting for surgery (Brunner) and in ours where they are on the surveillance programme. There is evidence that with increasing RMP there is increasing acetabular deficiency, volume and area[116]. Gose *et al* identified a posterior-superior component to the vector between the centroids of the acetabulum and femoral head in 82% of their cohort, they also reported a reduced acetabular volume in their cohort[46]. It is possible that a lack of acetabular coverage could lead to a greater sagittal plane component of hip displacement, and that a substantial element of anterior or posterior displacement may be predictive of future displacement to the point of requiring surgical intervention. The designers of any future surveillance programme may consider including routine measurements of anterior-posterior displacement to establish if there is an association with outcome.

We used RMP in this study because RMP is a faithful measurement of lateral hip displacement. We were able to combine measurements of RMP and FHPPR to

produce a vector diagram of displacement in the transverse plane. However, using two different imaging modalities is impractical and inefficient in the clinical practice.

The question remains as to whether it is possible to predict orthogonal measures of displacement from a single ultrasound assessment. From the *in vitro* study, we found that LHC was sensitive to posterior displacements of the hip. However, our clinical study showed good correlation between LHC and RMP in our sample population (cf. Chapter 5). It is likely that for most individuals hip displacement is dominated by lateral displacement, however we have identified a group who do have a significant anterior or posterior displacement. Linear regression was used to create an estimate of the lateral displacement using RMP as a faithful measure of lateral displacement and LHC and FHPPR as independent variable.

A multiple linear regression was calculated to estimate the lateral displacement of the femoral head from the two ultrasound measures LHC and FHPPR. A significant regression equation was found ($F(2,16) = 17.8$, $p<0.001$), with an R^2 of 0.68. The following equation can be used to predict the lateral component of the displacement, the output is similar to RMP in that it is an estimate of the proportion of the femoral head in the medio-lateral direction that is not covered by acetabulum (RMP_{US}), Equation 16.

$$RMP_{US} = 0.920 - 0.886\,LHC - 0.388\,FHPPR$$

Equation 16: Prediction of lateral displacement of femoral head from ultrasound measurements

The sample used in this study was smaller than that used in Chapter 5, to understand whether the inclusion of FHPPR improved the strength of the correlation the simple

linear regression was recalculated with LHC as the only independent variable to allow for comparison, this regression yielded an R^2 of 0.61.

These results show that the inclusion of FHPPR improves the predictive power of ultrasound-based measurements of the lateral component of hip displacement.

6.6.1. Clinical implications

The results of this study agree with those of Brunner et al[25] and Miller et al[26]. By visualising and measuring the position of the femoral head relative to the acetabulum in medio-lateral and anterior posterior directions, it has been established that there may be significant antero-posterior displacements in the population of children with CP with hip dysplasia. Current clinical practice is to only measure the lateral component of displacement, using RMP from planar radiography.

The progression of hip dysplasia and the mechanism of the femoral displacement is not fully understood. Like most secondary effects of cerebral palsy there is heterogeneity in the presentation of hip dysplasia and outcome of preventative and corrective interventions. The challenges associated with predicting the natural progression of the pathology within an individual as well as the likely efficacy of an intervention may in part be due to the limited understanding, and 3D modelling, of hip displacement.

It is common in clinical practice to categorise hips into different levels of risk for progressive hip dysplasia. There are some commonly used RMP thresholds which influence clinical decision making. These are listed in the table below (Table 18). The impact of 3D imaging can be appreciated in the following argument. If, the magnitude of the absolute displacement were found to be more informative than just the lateral

component, then in the cohort that we investigated two of the nineteen hips would have fallen into a higher risk category, (illustrated in Table 18, calculating the magnitude of the vectors displayed in Figure 47).

Clinical category	*RMP*	*Vector magnitude*
Normal (< 30%)	15	13
At risk (30% - 40%)	1	3
In need of intervention (> 40%)	3	3

Table 18: Categories of risk of progressive hip dysplasia when measuring RMP and the RMP/FHPPR vector magnitude

Brunner et al[25]'s work suggests that the channels through which the femoral head displaces are uni-directional. This study suggests that it may be possible to construct the vector of displacement from two orthogonal images, either from a single ultrasound assessment capturing images from orthogonal positions or by using X-ray to quantify the lateral displacement. Pragmatically, the ability to conduct a single assessment is superior however the abstract construction of the lateral estimate of displacement does present a further challenge for adoption. It is challenging to accept (and challenge) clinical measurements from images that cannot be easily verified 'by eye'. Further work is needed to establish if it is of significant clinical interest to routinely assess hip displacement in this way. If this were to be adopted, from a young age, it may be possible to detect those individuals with a significant anterior/posterior component to their displacement, even if the absolute magnitude of the displacement is not of immediate concern. Identifying this population and understanding how their hip displacement progresses may help to refine predictive equations and assign more accurate patient specific risk profiles in the future.

6.6.2. Limitations

There are several limitations to the methodology used in this study, for this reason the results should be interpreted as illustrative and not absolute. The main assumption, which underpins the data generated and displayed, is the assumption of laterality of the data. It was assumed that FHPPR is an orthogonal measure to RMP, that FHPPR is a pure sagittal plane measure, and RMP is a pure coronal plane measure. Whilst these assumptions do not reflect reality, if the results are interpreted as illustrative it is sufficiently robust to allow further investigation into the 3D directionality of the hip displacement. These results provide evidence of the ability for 3D ultrasound to detect and measure the direction of hip displacement in both the coronal and sagittal planes to indicate the need for further study. To remove the requirement to make this limiting assumption in a further study it would be necessary to create a global reference frame, where the probe position, and therefore the exact orientation and position of the ultrasound scans, is known. This could be achieved using motion capture and marker clusters both on the probe and in a fixed position on the child. Freehand 3D ultrasound is an established imaging method, particularly in the research domain[84–86,125], which utilises these concepts to construct volumetric, or 3D, ultrasound data from 2D image slices. The viability and reliability of the methodology is already largely proven[85–87].

The magnitude of each of the components of the vector is taken from each of FHPPR and RMP, it was assumed that these measures are equivalent and therefore no scaling factor was applied to either measurement. This assumption is based on the construction of each of the measures. Both measures are normalised by an estimate of the diameter of the femoral head. A similar assumption was tested in Chapter 5. The results showed that the estimates of femoral head diameter varied by approximately 10% when measuring femoral head diameter by different X-ray

methods and different ultrasound methods. An estimate of femoral head diameter routinely underpins clinical measurements of hip morphology.

This study also had a small sample size and did not have a full spectrum of hip displacement, in order to generalise the results further study over a greater sample size would be necessary. Finally, the reliability of the image acquisition and analysis has not been tested in this study. These analyses would be needed to fully understand the potential clinical utility of FHPPR.

6.7. Further investigation

Without further studies to validate the use of a 3D ultrasound assessment of the hip with multi-planar measurements to quantify the hip development the clinical implications remain speculative.

As discussed, without 3D clinical imaging on the study participants, it is not possible to verify the magnitude or scale of the impact of displacements that deviate from the lateral plane. However, given the results presented in this chapter, and other studies' results[25], it is clear that these displacements are present and reasonably prevalent (approximately 40% of cases).

It is plausible that in some individuals a genuine anterior or posterior displacement of the hip may be detected by the LHC measurement that would not have been detected by RMP from X-ray due to different imaging modalities underpinning the measurements.

6.8. Conclusions

This study explores the potential clinical utility of a measure of femoral head positioning in the sagittal plane. The results indicate the potential of the FHPPR index, used in conjunction with a measure of lateral displacement, to better describe the trajectory of hip migration. Further studies are needed to validate the clinical impact of this increased understanding of the hip positioning.

7. Review, recommendations, limitations and future work

7.1. Overview

This chapter summarises the preceding chapters and synthesises the findings from each of the studies. There were three separate workstreams documented in this thesis; a simulation, an *in vitro* experiment and a clinical study. Despite the written order of these studies, these were largely conducted in parallel, particularly the *in vitro* and *in vivo* work. Inevitably, there are some findings from the *in vitro* study that could have informed the design of the *in vivo* one, had these been conducted sequentially. Although this was not the ideal, it was necessary to commence the *in vivo* study early in the project to ensure sufficient numbers of subjects were recruited to the clinical study. Nevertheless, each study has contributed to our understanding of the measurement of hip migration and the design of hip surveillance programmes.

By the way of demonstration of the additional utility of 3D ultrasound imaging, this chapter concludes with two further investigations evaluating the potential of combining image volumes to visualise the hip in 3D. Exploration of the ultrasound image volumes allowed the morphologies of the two hips to be visualised using two techniques. Firstly by registering common bony landmarks from images taken from different perspectives to model the acetabulum border and femoral head in 3D. Secondly, a case study of two individuals with divergent presentations identified by the sagittal measurement of FHPPR, image volumes from these two hips were manually segmented to build 3D rendered models which can then be viewed from different planes.

7.2. Learnings from this thesis

7.2.1. Literature

Measuring hip dysplasia in cerebral palsy is a complex technical problem, and the factors that contribute to hip progression are difficult to resolve. Firstly, hip dysplasia arises secondary to a neurological injury which in itself varies in severity and phenotype. The damage to the developing brain results in alteration of the development of the musculoskeleton in a way that is not fully understood. There are challenges associated with quantifying and treating a symptom, and not the underlying pathology. Often, treatments are based on a plausible rationale but have little objective evidence to support them. For example, in the case of hip dysplasia, muscle imbalances around the hip are thought to be causative. Yet, the quantification of individual muscle forces around the hip has never been performed, and the routine proxy measures such as passive range of movement are ambiguous in their meaning and do not have predictive value. It is interesting to note that, in the development of Hermanson's predictive equation (CPUP score[29]) passive range of motion measurements were not included as independent variables in spite of being routinely collected.

Secondly, a flexible 3D hip geometry can be manipulated to produce different 2D images of the same underlying structures, as demonstrated by Reimer in his original work[27] (Figure 30), leading to ambiguity in the interpretation of the images and to erroneous measurements. The errors result from the relative rotations of the femoral and pelvic segments but also to the projection of these segments to the image plane. Where 3D images of the hip are acquired, they are still susceptible to differences due to the relative anatomical rotations of the femoral and pelvic segments. It follows that creating a simple to use, and clinically viable, index of the status of the complex 3D

problem that is the hip in CP is a challenge. In general, clinical measurements taken from 3D volumes remain 2D measures. The advantage of 3D over 2D imaging is that the plane of the measurement can be decided upon after image acquisition is completed and is therefore not susceptible to the same level of error in 2D images produced at the time of acquisition.

Finally, the point at which intervention is required, or hip dysplasia is diagnosed, is poorly defined. The current range of lateral hip displacements considered as threshold for intervention do not appear to be guided by stochastic or mechanical concerns. Rather, the threshold values seem to be informed by clinical experience. Less than 33% RMP is commonly referred to as ‘not at risk’ and above this 33% threshold hips are often classified as ‘at risk’. Typically hips at this point are pain free with sufficient range of movement to support the physical abilities of the individuals. The thresholds are defined as the point where it is perceived that the likelihood of progressive displacement resulting in dislocation is great enough to warrant the risks of treatment, even, in many cases, if there will be no immediate gain. Recently, a point of no return was identified by Wordie et al[42] which is significantly higher than the frequently quoted 33%. This study is exposed to the measurement uncertainty that surrounds RMP, the authors state that errors of approximately 10% are reported in the literature and they would be cautious of changes in measurement of less than 7%. If the tools for monitoring were improved, and an individual’s trajectory were better understood the risk and benefit profile of an intervention may change. Further, RMP is insensitive to out of plane displacements. The use of RMP to quantify hip dysplasia creates a bias in the development of treatments and understanding of the mechanisms. Research is often conducted where populations are classified based on the measured RMP, or RMP is used as the outcome measure to quantify the performance of an intervention.

The use of 2D imaging and the arbitrary assignment of thresholds to determine hip status also affects the quality of clinical research because reported outcomes are likely to be dependent on the threshold value and measurement uncertainty. The limitations of our practices may reinforce the currently held view, and may limit our ability to challenge the existing paradigm[126].

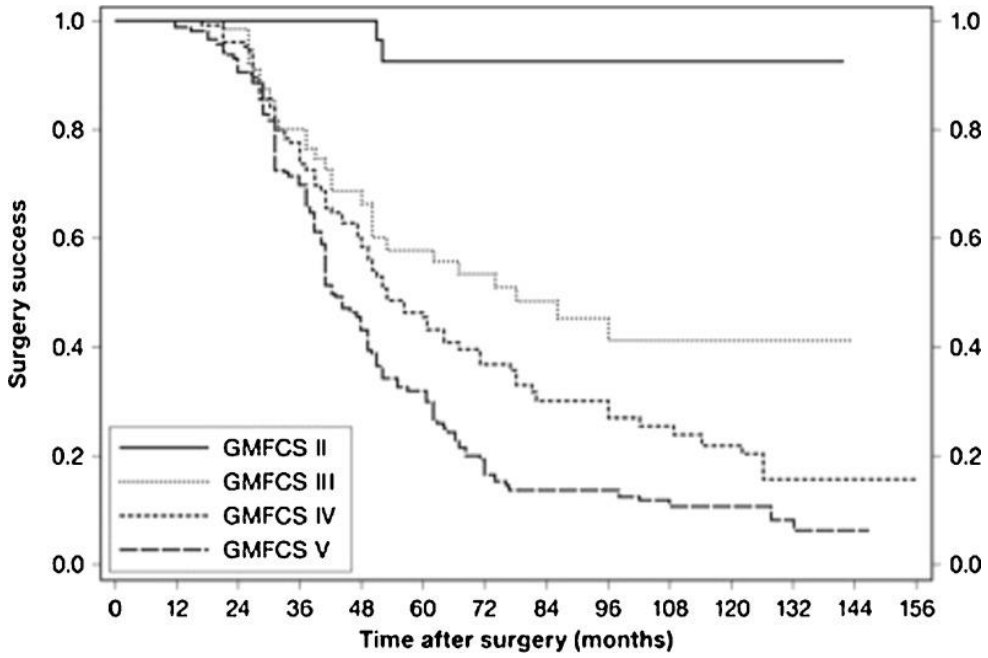


Figure 49: Survival plot showing the long term success rate from adductor surgery in the management of hip displacement in children with cerebral palsy. (Published by Shore et al[123])

Hip surveillance programmes were initially established approximately 30 years ago with the purpose of standardising and improving the care, and hip management of children with cerebral palsy. Across the globe, where these programmes have been adopted, dramatic reductions in hip dislocations rates have been reported. However there has also been an increase in preventative surgery rates reported. On introduction of a surveillance programme, Dobson *et al*[127] reported an increase in preventative surgeries (from 51% to 70.9%) and a decrease in reconstructive and salvage surgery (from 37.1% to 11.4% and 29% to 0% respectively). However, their results are only reported for a maximum of 4 years follow up period. In longer term

survival data from Shore et al[128] (Figure 49), the failure rate of preventative adductor releases increases significantly between 4 and 12 years (GMFCS V: failure rate of ~55% at 4 years which increased to over 90% at 12 years, GMFCS IV: failure rate of ~40% at 4 years which increased to over 80% at 12 years, GMFCS III: failure rate ~30% at 4 years which increased to ~60% at 12 years). In this study, failure was described as the need for subsequent surgery or an RMP> 50%. It is therefore likely that given a longer follow up period the reduction in reconstructive and salvage surgeries reported by Dobson et al[127] would be less dramatic. Dobson et al also reported a significant reduction in the average age of the patients at the time of their preventative surgeries. Given the relatively high failure rates at longer term follow up reported by Shore et al[128], the rational of preventative surgery, in the form of adductor releases, is less certain.

Preventative surgeries may not improve outcome above the natural history. Shore *et al*[128] defined failure as an RMP of greater than 50% or the requirement for further surgery. Comparing their rates to incidence of hip dysplasia in natural history studies (Table 19), it can be observed that the prevalence of hip dysplasia by GMFCS level is similar. There are methodological differences between the studies in Table 19 but we should be concerned by the similarity in the data reported.

Study	*MP limit*	*GMFCS I*	*GMFCS II*	*GMFCS III*	*GMFCS IV*	*GMFCS V*
Shore et al (at 12 years post adductor release)[128]	50 or further surgery	Not reported	10	60	80	90
Soo[33]	30	0	15	41	69	90
Connelley[129]	30	3	17	46	59	76
Hagglund[30]	33	5	13	50	62	68
Terjesen[32]	33	1	8	39	45	72

Table 19: Comparison of prevalence (%) of hip displacement by GMFCS level reported by different studies. RMP limit is the cut off limit used to categorise hips as displaced.

One of the significant challenges associated with identifying hips at risk of hip dislocation is the 'silent' nature of the hip displacement. Hip surveillance programmes facilitate the systematic identification of 'at risk' hips and recommend preventative interventions. 20 years after the introduction of routine hip surveillance in Sweden, Hagglund *et al*[24] reflected that the dislocation rates had fallen from 9% to 0%. However, to achieve this, 13% of individuals received preventative surgery and 44% of those required at least one repeated surgery. These results suggest that whilst hip surveillance programmes have caused a dramatic reduction in hip dislocation, the high number of preventative and repeated surgeries suggest that these programmes have poor specificity.

The work from Brunner *et al*[25] introduces the idea that the displacement may follow a uni-directional trajectory. Understanding the trajectory of the displacement may well be key to better differentiation and categorisation of individuals who require intervention to prevent symptomatic levels of hip dysplasia.

Hip surveillance programmes have had a positive impact on hip management in cerebral palsy, with dislocations now being very rare. However, there are opportunities to improve the performance and predictive value of these programmes at an individual level. Given the relatively poor long-term efficacy of standalone soft tissue surgeries, the presence of 'correction', and the identification of a point of no return at a relatively high RMP, perhaps consideration should be given to later bony surgeries in cases where dislocation, pain or deterioration in function are inevitable, or other factors indicating surgical intervention. However, in order to ensure that dislocation is

prevented whilst minimising overall rates of surgery we must first address the gap in our understanding of the progression and mechanism of this condition.

7.2.2. Simulation

The Monte Carlo simulation presented in Chapter 3 provides an insight into the possible misclassification of individuals as a consequence of measurement error in a typical hip surveillance programme. It is challenging to accurately model a clinical situation. Firstly, biological systems are highly complex. Secondly, it is impossible to include all the tacit understanding of the treating clinician gathered from their experiences and the expectations and fears of their patients and their families that might influence decision making in an explicit model. Our model showed agreement with the clinical outcomes reported by Soo *et al*[33] but at the level of the individual patient, if an intervention or treatment decision is made, it is impossible to understand what would have happened if that decision had not been made. This is the limitation of all statistically-based models.

Understanding the limitations of the methodology is key to interpreting the results of the simulation. There are some key trends and results that are plausible and would be hard to realise by conventional clinical (non-modelling) studies.

From our simulation we suspect the existence of a 'hidden' group of individuals, where the true progression of their hip dysplasia is slow, rendering them susceptible to misclassification of their hip status. This situation arises, when a clinician is presented with two X-rays from successive annual reviews. The calculated RMPs suggest that there has been a progression in the displacement of the hip, however this perceived increase in RMP has actually arisen from a change in position of the individual at the point of X-ray acquisition, a change in contrast or parameters of the acquired X-ray or

simply a difference in the identification of the bony landmarks and calculation of RMP. However, the progression that is measured is interpreted as a cause for concern, and the individual may well be referred for orthopaedic treatment.

Our model suggests that in certain populations, a positive indication for intervention is more likely to arise from measurement error, than from a genuine progression. This tendency is off-set by increasing the time period between the X-rays, increasing the likelihood and magnitude of true progression of the hip displacement between the assessments. This trend is seen at all GMFCS levels, however it is the most functionally able, GMFCS level III, who are most at risk of a mis-classification and receive the greatest benefits from prolonging the time between X-rays. The structure of hip surveillance programmes, with regular review, mitigates against the risks associated with missing an individual who is at risk. Intervention thresholds are largely defined to facilitate preventative interventions, prior to the hips becoming symptomatic, as such the risk of declining function or an increase in discomfort through misclassification is low. This is also mitigated further by including extra “out-of-programme” reviews for in the case of the individual with symptom progression.

Finally, the accuracy of RMP measurement is increased by taking repeated radiographs at a relatively short time interval. However, there is the complex interaction between the benefits derived from more accurate measurement and identification of “at risk” individuals against the increased exposure to ionising radiation from the X-rays for all individuals in the programme and increased resource required to acquire and interpret multiple images.

These results highlight a group of individuals who may benefit significantly from better identification and differentiation of those at risk of progressive hip dysplasia and those

whose hips will either remain stable, or stabilise at an acceptable functioning and pain free level of hip dysplasia. A more stratified approach to the design of hip surveillance programmes may be advantageous, however it is unlikely that significant improvements will be made without improving the assessment tools used in these programmes.

7.2.3. From the *in vitro* work

The complexities of validating a novel clinical index or measurement have been discussed, particularly in situations where the clinical standard cannot be considered to be a gold standard. The purpose of the *in vitro* study was twofold. Firstly, to identify potential indices to quantify displacements in the both the sagittal and coronal planes. Secondly, to systematically assess the performance of the indices and the repeatability of the measurement of the indices. Two hip phantoms were 3D printed, from a CT scan of a child with cerebral palsy (GFMCS V). A custom mechanical rig was designed to manipulate the hips into known positions. Ultrasound images were acquired with the hips in different positions to allow for indices to be derived that were sensitive to the genuine displacements and insensitive to anatomical rotations and out of plane displacements.

Two novel indices were developed. The lateral head coverage (LHC), derived from 3D volumes acquired with the hip in a side-lying position, with the probe held over the lateral aspect of the femoral head, and femoral head posterior position ratio (FHPPR) which was derived from image volumes acquired in a supine position. Both measures were constructed in a similar way, with the slice chosen for analysis selected by identifying the slice of maximal cross-sectional area of the femoral head in both the

sagittal and coronal plane images. The indices are then simple ratios computed from two linear measurements within a single ultrasound slice.

Both indices showed excellent sensitivity to genuine displacements in the plane of the measurement (coronal for LHC and sagittal for FHPPR) and both were relatively insensitive to changes in the relative rotation of the femoral and pelvic segments, with one exception. LHC was found to be sensitive to ab/adduction, changing by 1% for each 2° change in ab/adduction (Table 11). LHC was also found to be highly sensitive to posterior displacement. This result is explained by the morphology of the acetabulum. Whilst this sensitivity does result in some greater deviations in the agreement between X-ray and ultrasonic measurement of lateral displacement, it does provide reassurance that LHC detects displacement irrespective of trajectory. In contrast, the FHPPR index was found to be insensitive to lateral displacements.

7.2.4. From the in vivo work

There were three key clinical studies described in this thesis. The first is a “levels of agreement” study, comparing LHC and RMP in a clinical population. Despite RMP not being considered to be a gold standard measurement, it is the current clinical standard, and it was important to compare these indices for the sake of establishing concurrent validity. The second study was a repeatability study evaluating the inter- and intra- assessor reliability of measuring LHC from ultrasound volumes. The third study was more explorative, and demonstrated the potential for a more extensive ultrasound assessment of the hip allowing the quantification of the displacement in both the coronal and sagittal planes to be visualised. The first two studies produced

encouraging results, showing a strong correlation between LHC and RMP, indicating that LHC was detecting the lateral displacement measured by RMP.

The final clinical study investigated the feasibility of quantifying, and the potential prevalence of anterior-posterior displacement. In this study, we observed that there was a degree of non-lateral displacement greater than ±13° from the pure lateral direction in nearly half of the subjects. The magnitude and direction of the displacement was variable, with the most extreme cases demonstrating a trajectory of displacement 30° posterior to the pure lateral direction. By including the A-P measurement of hip displacement (FHPPR) in a linear regression we could improve the agreement between an ultrasonically derived index of lateral hip displacement and RMP. This result suggests that ultrasonic investigations may be used to chart the trajectory of hip displacement in the transverse plane.

7.3. Clinical implications

Combining our learnings from the literature with the results of the studies documented in this thesis some potential clinical impact can be anticipated. There are also several limitations which limit the confidence in the results and warrant further study. A key theme identified is the discrepancy between population and personalised medicine. Clinical trial results inform us of population-level responses to treatment. Treatment strategies and clinical pathways, such as those used in hip management, are designed and adopted, according to these findings. However, for some cohorts the sensitivity and specificity limit their efficacy, and predicting the likelihood of a successful outcome for some individuals is no easier than predicting the outcome from a coin toss.

7.3.1. Improving the rational and evidence-base for intervention

Currently, hip surveillance programmes group individuals into broad categories based on their gross motor function and in some cases the diagnostic phenotype. However, an individual with cerebral palsy may have particular characteristics which may more or less predispose them to a positive (or a negative) outcome from treatment.

Muscle imbalance around the hip is thought to play an important role in the development of hip dysplasia, in particular the hip adductor muscles. However, the literature reports variable success from adductor releases as a preventative procedure (where some success rates are reported to be about 50%[128]). Let us consider three plausible reasons why this variability in outcome may occur. Firstly, the surgeries may not be too conservative, the abnormal loading forces are not relieved, and the muscle imbalance remains. Secondly, the cause of the altered adductor function is not addressed so that the problem may continue to advance. Finally, tightness of the hip adductors may be only one of many abnormal features that drive hip dysplasia.

Typically, at follow up, to assess the outcome and success of adductor surgery, a hip assessment would be conducted. This may include a physical examination focusing on passive range of motion, pain scores and imaging to assess the hip positioning. These assessments do not assess the contributions of individual hip adductors pre- and post-surgery to the forces acting on the hip. Given the variable outcomes from adductor surgery it is likely that our rational for lengthening the adductors is incomplete or flawed. For example, Larkin-Kaiser *et al*[130] demonstrated a strong association between resting sarcomere length and titin weight in the gracilis, and hip migration percentage in a group of children with spastic cerebral palsy. While it may not be possible to repeatedly obtain biopsy samples, in a typical clinical setting, measurements from imaging (for example, ultrasound[85]) should be acquired that allow

for direct observation of the adductor structure and function and hip position before surgery, as an immediate result of the treatment, as well as surveillance of these outcomes in the longer term. It is conceivable, that one, or multiple, of these factors would be identified that correlate with positive long-term outcomes, and a score developed that was predictive of long-term hip instability. In addition, more detailed investigations of muscle structure, composition and function may elucidate the mechanisms of hip displacement leading to better treatments of hip displacement.

7.3.2. Soft-tissue features

The congruence of the femoral head and the acetabulum is not the only structural feature that may lend the hip stability. There are soft tissue structures around the hip that may contain the hip mechanically, and may be altered in children with cerebral palsy and other developmental conditions.

X-ray does not allow for the visualisation of the soft-tissues. Historically measurements of muscle volumes required MR imaging however, Passmore et al[87], Barber *et al*[86] and Noble *et al*[131] have shown that 3D ultrasound can reliably estimate the volume of muscles in children with cerebral palsy. Vanmechelen *et al*[132] have shown that accurate estimates of lower limb muscle volumes can be made from measurements of maximum cross-sectional area and muscle belly length. Their technique would allow the assessment of muscles, such as the hip adductors, where direct measurements of muscle volume may be practically difficult in a clinical setting. Deficits in muscle volume have been reported in most of the major lower limb muscle groups in children with cerebral palsy. Additionally, changes in composition of the muscles of cerebral palsy have been demonstrated and detected by ultrasound[133]. Multani *et al*[134], recently

called for routine monitoring of muscle volumes in children with cerebral palsy undergoing botulinum toxin type A injections for therapeutic management of spasticity, amongst concerns of the detrimental effect of botulinum toxin on muscle volume and strength. These same concerns may apply to the muscle releases and other surgeries performed in the management of the hip.

Connective tissues are known to play a significant part in hip stability and mechanics, particularly in the dysplastic hip (DDH)[135]. The acetabular labrum provides a dense fibrocartilaginous ring that effectively increases the depth of the acetabulum. Horii *et al*[136] looked at the development of the acetabulum and acetabular labrum in the normal child using MRI. They divided individuals into groups dependent on age. Their results showed that younger children (aged 6-11years) had less acetabulum coverage than children over 12 and adults, particularly postero-superior coverage. However, the acetabular labrum coverage showed an inversely proportional relationship to age, with significantly greater coverage seen across the whole labrum in the 6-11 years and the greatest differences seen in the postero-superior aspect. In the normal adult the labrum is thought to play a significant role in the joint stability and function, providing the seal to ensure efficient movement through the trapping of a pressurized fluid film distributing pressures in the joint and preserving the cartilage[137]. The labrum is not thought to play a significant role in load support in the normal hip, studies suggest it is responsible for 1-2% of the load support, however in the dysplastic hip it is thought to provide 4-11% of the support of the load[137]. These findings indicate the potential structural importance of the acetabular labrum, particularly in the young. There have been several studies that evaluate the acetabular labrum under ultrasound, these studies focus on an adult population and centre around labral tears[119,120]. To our

knowledge, in depth analysis of the role of the labrum in individuals with CP have not been conducted.

The interaction between muscle function, size and strength and bony development are beginning to be explored[138,139]. Although an asymmetry in forces generated by muscles about the hip is thought to be at least partially causative of hip displacement[26,27], local muscle thickness and composition appears to affect the local apposition of bone. It follows that local muscular morphology may influence the development of the bony structures that make up the hip. Asymmetry in muscle volume as well as muscle tightness and dysfunction may, in part, explain progressive displacement and the trajectory of that displacement.

7.4. Future work

7.4.1. Can we identify predictive factors to better tailor hip management programmes for children with cerebral palsy?

This thesis introduced the concept of mathematical simulation to help design hip surveillance programmes and understand their impact on treatment decision making. Also described, were a novel method and indices for assessing hip position in children with cerebral palsy. However, there are unanswered questions that remain in need of further investigation. These future studies will be required before advances in the clinical management of the hip in CP are realised.

The potential for understanding the significance of the trajectory of displacement from 3D ultrasound images requires two key adaptations to the clinical study design described in this thesis. Firstly, the ultrasound methodology should be validated against a gold standard 3D imaging modality such as CT or MRI. Secondly, it may be

necessary to track the ultrasound probe relative to the pelvis to ensure that the positioning and orientation of the image volumes is known. This would allow image volumes obtained from different perspectives to be 'stitched together' or merged. This is particularly important considering the views that are available under ultrasound imaging. In ultrasound, particularly when imaging areas with bony structures, it is not possible to visualise surfaces that sit deeper than the most superficial bony surface. In order to build up a more complete image it is beneficial to move the probe around the area and combine the image volumes. The tracking could be achieved using an optical motion capture system, similar to that described in others studies[85].

To investigate the potential value of superimposing ultrasound volumes an investigation was conducted using one of the subject's images. Ultrasound volumes acquired from both the lateral and anterior skin surfaces about the same hip were analysed together. Instead of registering these volumes using motion capture, common bony landmarks were identified in each volume with the assumption that the image volumes were collected from perpendicular perspectives. The slice of maximal cross-sectional area of the femoral head was identified and the sphere of best fit plotted. The acetabulum was then segmented manually using both coronal and sagittal planes. The coordinates of the centroid of the estimated spherical femoral head and the radius were recorded. The coordinates along the inferior border of the acetabulum were identified and recorded from each of the lateral and anterior rendered images. The coordinates describing the femoral head estimates were then plotted over each other by transposing the coordinates for one of the image volumes to the other image volume. The radius of the femoral head was taken to be the mean radius from the two volumes. The coordinates of the lateral aspect of the acetabulum from both image volumes were then transposed and plotted (Figure 50).

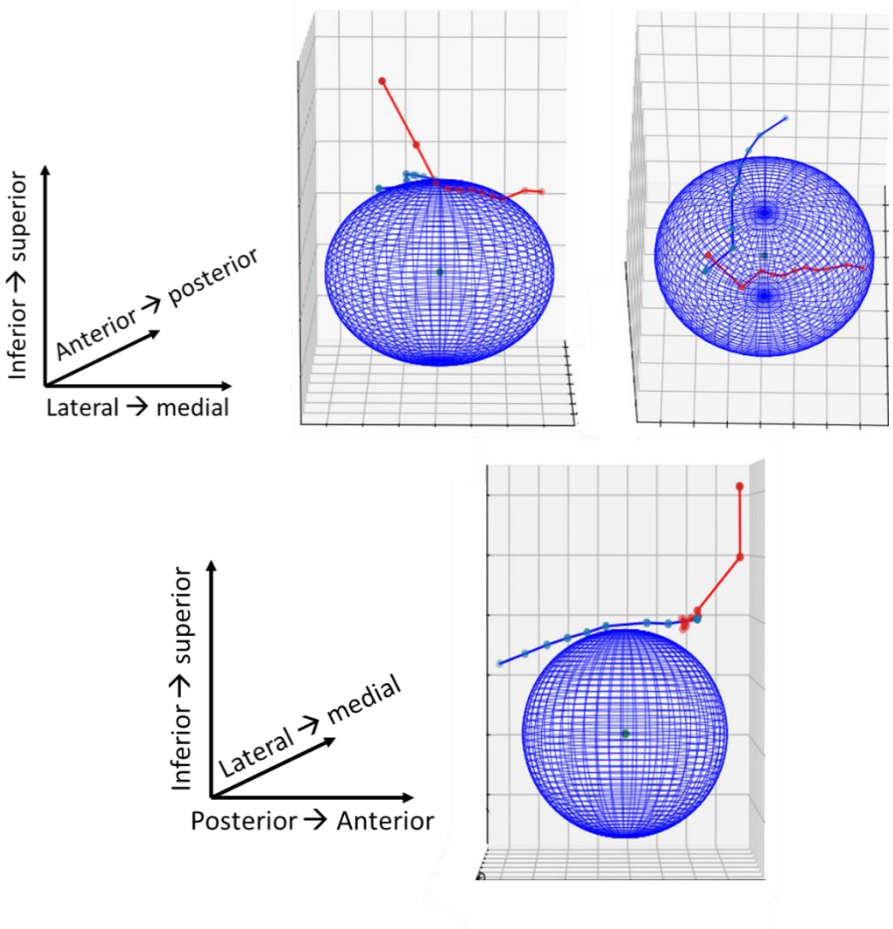


Figure 50: Image displaying a reconstruction of the hip from coordinates taken from 2 ultrasound volumes. The coordinates have been transposed and superimposed on each other to create a 3D visual of the acetabular border and femoral head. The blue line are the transposed coordinates of the inferior border of the acetabulum from the laterally acquired ultrasound image volume. The red line is the anterior border of the acetabulum acquired from the image volumes acquired in supine with the probe over the anterior aspect of the hip.

Figure 50 shows the estimated femoral head, from both image acquisitions. The centroids are shown, plotted over each other. The blue points and line are the transposed coordinates of the inferior border of the acetabulum from the image volume acquired over that lateral aspect of the hip, and the red points and line are from the volume acquired from the anterior aspect of the hip. The initial investigation into the value of this approach shows the potential to assess the femoral head coverage by the shadow projected from the acetabulum over the femoral head. Comparing Figure 50 to Figure 51, the method appears feasible but would be improved by knowing the relative orientation of the two ultrasound image volumes.

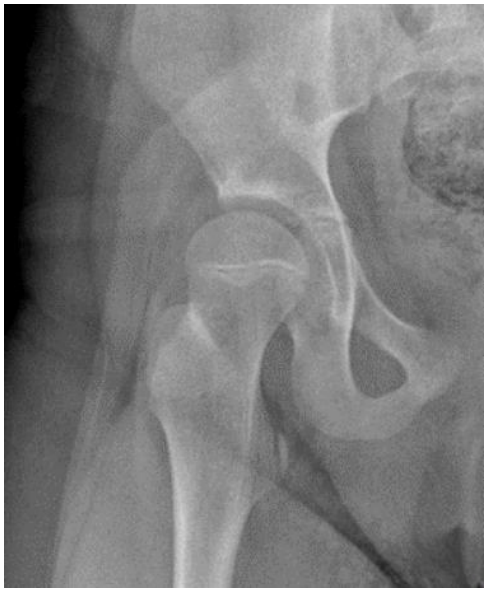


Figure 51: X-ray of the Recruit 15 right hip, selected for the further analysis.

To validate the use of 3D ultrasound for visualising the bony structures of the hip, CT scans would typically be required. However, if MRI scans were acquired instead it would be possible to assess the performance of the ultrasound in identifying and measuring some of the soft tissue structures. Of particular interest would be the hip adductor muscles, namely adductor longus and gracilis as these are most commonly targeted in preventative surgeries, as well as their antagonists. The acetabular labrum is considered key in hip stability in children with developmental dysplasia, understanding how the labrum contributes to hip stability in children with cerebral palsy, and whether changes in the labrum impact the trajectory of hip displacement would be valuable when searching for explanatory features.

7.4.2. 3D rendering of ultrasound volumes – a case study

The clinical studies described in Chapters 5&6 have discussed the development and preliminary validation of indices to describe the displacement of the hip. However, clinicians often use ultrasound for qualitative assessment. There may be a role for

more qualitative interpretation of hip images, allowing clinicians to visualise the hip structures in 3D. To illustrate, two case studies were selected with divergent suspected displacements in the anterior-posterior direction. On review, the measured LHC for these hips was the same (62%), the RMPs for the two hips were 28% and 52%, and the calculated lateral component of displacement (RMP_{US}), from ultrasound assessment, were 31% and 43% respectively. Full details are displayed in Table 20.

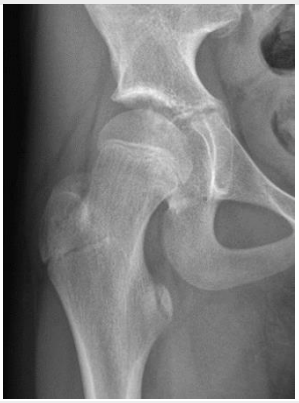
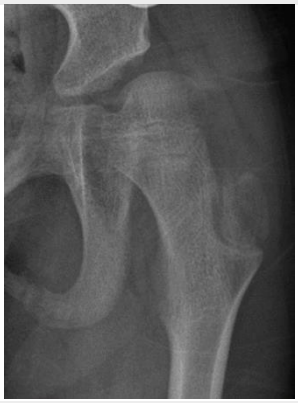
	Hip 1	*Hip 2*
Gender	Female	Male
Age	13.25	7.75
Side	Right	Left
RMP (measured)	28%	52%
LHC(measured)	62%	62%
FHPPR (measured)	16%	-16%
RMP_{US} (calculated)	31%	43%
X-ray		
Reason for investigation	Suspected posterior displacement	Suspected anterior displacement

Table 20: Descriptions and X-rays of the two hips that underwent further evaluation

The hip demonstrating posterior positioning relative to the anterior aspect of the acetabulum was the right hip of a 13-year-old girl, and the hip with anterior positioning relative to the anterior aspect of the acetabulum was the left hip of a 7-year-old boy (Table 20). For both hips, volumes acquired from both side lying and supine positions were manually segmented and rendered using the segment editor module from Slicer version 4.10.1. From each volume, the renders were manipulated to be visualised in

both coronal and sagittal planes from each of the image volumes acquired for each hip (Figure 53).

Due to the physics of imaging ultrasound, the sagittal (or coronal plane) views of the hip taken from different image volumes will not appear equivalent. Bony surfaces prevent deeper structures being visualised, so one must be careful of their interpretation. The renders in Figure 52 a and b were created from two ultrasound volumes, the top row of each panel from a scan taken in side-lying. It is from the coronal plane slice from this scan that LHC was measured. The bottom row of each panel was from an image volume, of the same hip acquired in supine lying. The renders from each of the volumes are displayed in both the coronal and sagittal planes. The yellow and red/brown render represents the portions of the acetabulum and femoral head respectively, that can be visualised directly from the ultrasound volumes. The renders have been smoothed to remove slice to slice variability but have not otherwise been modified. The green spheres are the best fit spheres for the femoral head. The spheres have been integrated into the images to aid in interpretation. The coronal and sagittal views from the two different image volumes have been displayed above and below each other, this is to aid interpretation and visualisation. By imagining superimposing the coronal images on top of each other, the profile of the acetabulum relative to the femoral head begins to be visualised. It can also be seen that this looks similar to what would be expected from viewing the Hip 1 X-ray (Table 20). The same process with the sagittal plane images allows for the visualisation of the acetabulum and femoral head in the sagittal plane.

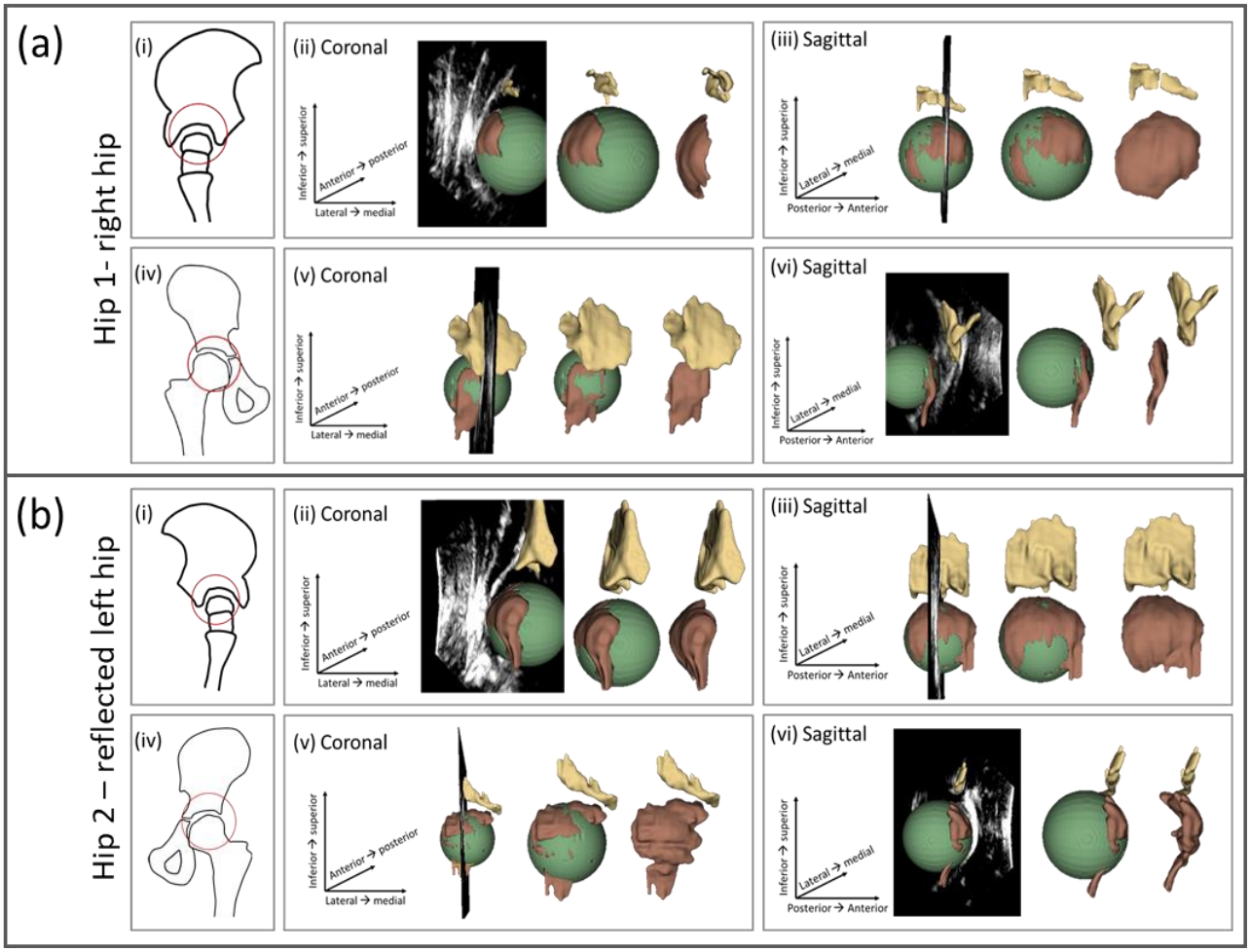


Figure 52: Panel (a) shows the 3D renders from hip 1 (right hip), panel (b) shows the 3D renders of hip 2 (left hip), for ease of comparison the left hip images have been reflected. Image volumes acquired in both side lying and supine lying. (i) schematic of acquired volume in side lying, (ii) coronal plane views of render from side lying data, (iii) sagittal plane views of the side lying render. (iv) schematic of acquired volume in supine, (v) coronal plane views of render from supine data, (iii) sagittal plane views of the supine data render. Key: red/brown render = viewable portion of the femoral head, yellow render = acetabulum segment, green render = estimated femoral head.

Panel (b) displays the same set of images for Hip 2 however as the hip is a left hip, the renders have been reflected to allow for comparison with Hip 1 (Figure 52, Panel a). Comparing the surface renders, there is a clear difference in the position of the femoral head relative to the acetabulum in the sagittal plane (a view that would not be visible from the conventional anterior-posterior X-ray imaging used in hip surveillance programmes). From Figure 52 Panel a (vi), Hip 1, a large gap is observed between the anterior acetabular border and the anterior surface of the femoral head, indicating that the hip is positioned posteriorly. Figure 52 Panel b (vi), Hip 2, shows the anterior

surface of the femoral head protruding anteriorly beyond the anterior border of the acetabulum, indicating that the femoral head is relatively anteriorly positioned.

From the X-rays of these two hips (Table 20), Hip 1 has an RMP of 28%, which means the hip would be categorised as ‘not at risk’. Hip 2 has an RMP of 52%, indicating that the hip is showing significant displacement, and would likely be indicated for orthopaedic intervention. Values from the modified ultrasound index (RMPus) suggest a similar level of lateral displacement for these hips. However, the ultrasound index (FHPPR) indicates displacements in the antero-posterior direction consistent with the qualitative images in Figure 52.

At present, we do not know whether the presence of posterior or anterior displacement of the hip is a significant factor in the progression of hip dysplasia, or whether measurements of lateral displacement are sufficient to predict progression. In the case of Hip 1, ultrasound assessment, as described, may have altered the clinician’s view of the significance of the hip position.

7.4.3. Opportunities for 3D ultrasound in hip surveillance

Ultrasound is a common and familiar imaging method in the hospital setting. It is a trusted and safe imaging technique that is widely accepted as a standard non-invasive technique for many clinical assessments including foetal, abdominal and vascular investigations. The potential for 3D ultrasound to provide multi-planar measurement of hip displacement has been demonstrated and discussed in this thesis however ultrasound also provides the opportunity for taking multiple images at the same assessment. Averaging measurements from multiple images increases confidence in the accuracy of the measurement. However, the greater impact may come from imaging the hips in different positions. Current clinical practice aims to reduce the

variability in hip position by standardising position protocols, increasing the reliability of the measurement of RMP, however there is likely to be value in assessing the hip in different positions. In a CT study conducted by Chung et al[140] the hips of children with CP were imaged in two positions, at rest and in a hamstring stretch. They observed a posterior shift in the position of the femoral head when individuals were placed in a hamstring stretch. The magnitude of this displacement was significantly greater in the group with RMP greater than 30% compared to those with RMP less than 30%. Posterior displacement was measured as a percentage of femoral epiphysis diameter, the average displacement in the group with RMP greater than 30% was 7.4% compared to just 0.5% for the group with RMP below 30%. The index of anterior-posterior displacement developed in this thesis (FHPPR) may be sensitive to the instability of the femoral head under this manoeuvre.

Ultrasound would allow for a more personalised approach to surveillance programmes without consideration of the risk from repeated exposure to ionising radiation. For example, for the child classified as GMFCS II or III, with low likelihood of reaching significant critical levels of hip displacement; instead of performing preventative surgeries at an early presentation, the child could be placed on a monitoring programme with frequent repeated assessments using 3D ultrasound to ensure that excessive progression is not missed.

Translation of a new measurement method from the research domain to clinical practice is fraught with difficulty, especially where there exists a well-established method. Further, the requirements to upskill the clinical population to acquire, analyse and understand the new assessment requires significant resource and is a barrier to adoption. The familiarity and prevalence of ultrasound assessment within routine clinical practice minimises the cultural changes required to implement ultrasound in

hip surveillance programmes, however there are still significant challenges to overcome. New techniques and measurements may be criticised for their relatively small evidence bases and lack of historical data. The development of the RMP index preceded the adoption of routine hip surveillance programmes, and the improvements in outcome realised from the structured programmes validates the further use of RMP. The introduction of a novel technique is unlikely to gain traction amongst the clinical community until its efficacy is comprehensively demonstrated. In the near future, the most likely application of the ultrasound method described is to serve as an adjunct to existing radiographic investigations, where further characterisation and visualisation of the 3D geometry is sought.

8. References

1. Bax, M. . Terminology and Classification of Cerebral Palsy. *Dev. Med. Child Neurol.* **6**, 295–97 (1964).

2. Musielak, B., Idzior, M. & Józwiak, M. Evolution of the term and definition of dysplasia of the hip -a review of the literature. *Arch. Med. Sci.* **11**, 1052–1057 (2015).

3. Van Naarden Braun, K. *et al.* Birth prevalence of cerebral palsy: A population-based study. *Pediatrics* **137**, 1–9 (2016).

4. Stavsky, M. *et al.* Cerebral palsy-trends in epidemiology and recent development in prenatal mechanisms of disease, treatment, and prevention. *Front. Pediatr.* **5**, 1–10 (2017).

5. Cans, C. *et al.* Prevalence and characteristics of children with cerebral palsy in Europe. *Dev. Med. Child Neurol.* **44**, 633–640 (2002).

6. Surman, G. *et al.* UKCP: A collaborative network of cerebral palsy registers in the United Kingdom. *J. Public Health (Bangkok).* **28**, 148–156 (2006).

7. Novak, I. *et al.* A systematic review of interventions for children with cerebral palsy: State of the evidence. *Dev. Med. Child Neurol.* **55**, 885–910 (2013).

8. Sewell, M., Eastwood, D. M. & Wimalasundera, N. Managing common symptoms of cerebral palsy in children. *BMJ* **349**, (2014).

9. Palisano, R. *et al.* Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev. Med. Child Neurol.* **39**, 214–223 (1997).

10. Graham, K. H. Classifying cerebral palsy. *J. Pediatr. Orthop.* **25**, 127–128 (2005).

11. Alriksson-Schmidt, A. & Hägglund, G. Pain in children and adolescents with cerebral palsy: a population-based registry study. *Acta Paediatr.* **105**, 665–670 (2016).

12. Eriksson, E., Hägglund, G. & Alriksson-Schmidt, A. I. Pain in children and adolescents with cerebral palsy - A cross-sectional register study of 3545 individuals. *BMC Neurol.* **20**, 1–9 (2020).

13. Parkinson, K. N., Dickinson, H. O., Arnaud, C., Lyons, A. & Colver, A. Pain in young people aged 13 to 17 years with cerebral palsy: Cross-sectional, multicentre European study. *Arch. Dis. Child.* **98**, 434–440 (2013).

14. Bagg, M. R., Farber, J. & Miller, F. Long-term follow-up of hip subluxation in cerebral palsy patients. *Journal of pediatric orthopedics* vol. 13 32–36 (1993).

15. McLaughlin, J. F. The measurement of health-related quality of life (HRQL) in children with cerebral palsy. *Eur. J. Neurol.* **8**, 183–193 (2001).

16. Wake, M., Salmon, L. & Reddihough, D. Health status of Australia children with mild to severe cerebral palsy: Cross-sectional survey using the Child Health Questionnaire. *Dev. Med. Child Neurol.* **45**, 194–199 (2003).

17. McCarthy, M. L. *et al.* Comparing reliability and validity of pediatric instruments for measuring health and well-being of children with spastic cerebral palsy. *Dev. Med. Child Neurol.* **44**, 468–476 (2002).

18. Vargus-Adams, J. Health-related quality of life in childhood cerebral palsy. *Arch. Phys. Med. Rehabil.* **86**, 940–945 (2005).

19. Landgraf, J. M. Child Health Questionnaire (CHQ). in *Encyclopedia of Quality of Life and Well-Being Research* (ed. Michalos, A. C.) 698–702 (Springer Netherlands, 2014). doi:10.1007/978-94-007-0753-5_324.

20. Houglum, P. A. & Bertoti, D. B. *Brunnstrom's Clinical Kinesiology*. (F.A. Davis Company,

2012).

21. Fairhurst, C., Turner, S., Lundy, C. & Norman-Taylor, F. The hip in cerebral palsy. *Paediatr. Child Health (Oxford).* **24**, 208–214 (2013).

22. Osborne, D., Effmann, E., Broda, K. & Harrelson, J. The development of the upper end of the femur, with special reference to its internal architecture. *Radiology* **137**, 71–76 (1980).

23. Cooke, P. H., Cole, W. G. & Carey, R. P. L. Dislocation of the Hip in Cerebral Palsy - Natural History and Predictability. *J. Bone Jt. Surg.* **71**, 441–446 (1989).

24. Hagglund, G., Alriksson-Schmidt, A., Rodby-Bousquet, E., Wagner, P. & Westbom, L. Prevention of dislocation of the hip in children with cerebral palsy: 20 year results of a population based prevention programme. *Bone Jt. J.* **96**, 1546–1552 (2014).

25. Brunner, R., Ricard, C. & Robb, J. Morphology of the acetabulum in hip dislocations caused by Cerebral Palsy. *J. Pediatr. Orthop. B* **6**, 207–211 (1997).

26. Miller, F., Slomczykowski, M., Cope, R. & Lipton, G. E. Computer modeling of the pathomechanics of spastic hip dislocation in children. *J. Pediatr. Orthop.* **19**, 486–492 (1999).

27. Reimers, J. *The stability of the hip in children. A radiological study of the results of muscle surgery in cerebral palsy. Acta orthopaedica Scandinavica. Supplementum* vol. 184 (1980).

28. Miller, F. & Bagg, M. R. Age and migration percentage as risk factors for progression in spastic hip disease. *Developmental Medicine and Child Neurology* vol. 37 449–455 (1995).

29. Hermanson, M., Hägglund, G., Riad, J., Rodby-Bousquet, E. & Wagner, P. Prediction of hip displacement in children with cerebral palsy DEVELOPMENT OF THE CPUP HIP SCORE. **97**, 1441–1444 (2015).

30. Hägglund, G., Lauge-Pedersen, H. & Wagner, P. Characteristics of children with hip displacement in cerebral palsy. *BMC Musculoskelet. Disord.* **8**, 101 (2007).

31. Larnert, P., Risto, O., H??gglund, G. & Wagner, P. Hip displacement in relation to age and gross motor function in children with cerebral palsy. *J. Child. Orthop.* **8**, 129–134 (2014).

32. Terjesen, T. The natural history of hip development in cerebral palsy. *Dev. Med. Child Neurol.* **54**, 951–957 (2012).

33. Soo, B. *et al.* Hip Displacement in Cerebral Palsy. *J. Bone Jt. Surg.* **88**, 121 (2006).

34. Scrutton, D. & Baird, G. Surveillance measures of the hips of children with bilateral cerebral palsy. *Arch. Dis. Child.* **76**, 381–4 (1997).

35. Shore, B. J. & Graham, H. K. Management of Moderate to Severe Hip Displacement in Nonambulatory Children with Cerebral Palsy. *JBJS Rev.* **5**, e4 (2017).

36. Park, J. Y. *et al.* Progression of Hip Displacement during Radiographic Surveillance in Patients with Cerebral Palsy. *J Korean Med Sci* **31**, 1143–1149 (2016).

37. Hägglund, G., Lauge-Pedersen, H. & Persson, M. Radiographic threshold values for hip screening in cerebral palsy. *J. Child. Orthop.* **1**, 43–47 (2007).

38. NHS Scotland. Cerebral Palsy Integrated Pathway Scotland (CPIPS). https://apcp.csp.org.uk/system/files/cpips_manual_2017.pdf (2017).

39. O'Donnell, M. *et al. Hip surveillance - Bottom Line 'Evidence-Informed' Recommendations for the Hip Surveillance in Individuals with Cerebral Palsy*. (2016).

40. Kleinberg, S. & Leiberman, H. The Acetabular Index in infants in relation to congenital dislocation of the hip. *Arch. Surg.* **32**, 1049–1054 (1936).

41. WAYNE 0. SOUTHWICK, M.D.t, NEW HAVEN, C. F. Osteotomy Through Lesser Trochanter for Scfe. *Surgery* (1967).

42. Wordie, S. J., Bugler, K. E., Bessell, P. R., Robb, J. E. & Gaston, M. S. Hip displacement in

children with cerebral palsy. *Bone Joint J.* **103**-**B**, 411–414 (2021).

43. Scrutton, D., Baird, G. & Sneeton, N. Hip dysplasia in bilateral cerebral palsy : incidence and natural history in children aged 18 months to 5 years. *Dev. Med. Child Neurol.* **43**, 586–600 (2001).

44. Hägglund, G. *et al.* Prevention of dislocation of the hip in children with cerebral palsy. The first ten years of a population-based prevention programme. *J. Bone Joint Surg. Br.* **87**, 95–101 (2005).

45. Chung, C. Y., Park, M. S., Choi, I. H., Cho, T. & Lee, K. M. Morphometric analysis of acetabular dysplasia in cerebral palsy. *J. Bone Jt. Surg.* **88**, 243–247 (2006).

46. Gose, S. *et al.* Morphometric Analysis of Acetabular Dysplasia in Cerebral Palsy: Three-dimensional CT Study. *J Pediatr Orthop* **29**, 896–902 (2009).

47. Chung, M. K. I. *et al.* Functional status and amount of hip displacement independently affect acetabular dysplasia in cerebral palsy. *Dev. Med. Child Neurol.* **59**, 743–749 (2017).

48. Park, M. S. *et al.* Two-Dimensional Computed Tomographic Measurement of Acetabulum V Reliability , Validity , and Limitation. *J Pediatr Orthop* **28**, 812–818 (2008).

49. Robin, J. *et al.* A classification system for hip disease in cerebral palsy. *Dev. Med. Child Neurol.* **51**, 183–192 (2009).

50. Gose, S. *et al.* Verification of the Robin and Graham classification system of hip disease in cerebral palsy using three-dimensional computed tomography. *Dev. Med. Child Neurol.* **53**, 1107–1112 (2011).

51. Scrutton, D. & Baird, G. Surveillance measures of the hips of children with bilateral cerebral palsy. *Arch. Dis. Child.* **76**, 381–384 (1997).

52. Pountney, T., Mandy, A., Green, E. & Gard, P. Management of hip dislocation with postural management. *Child. Care. Health Dev.* **28**, 179–185 (2002).

53. Graham, H. K. *et al.* Does Botulinum Toxin A Combined with Bracing Prevent Hip Displacement in Children with Cerebral Palsy and “Hips at Risk”? *J. Bone Jt. Surgery-American Vol.* **90**, 23–33 (2008).

54. Boyd, R. N. *et al.* The effect of botulinum toxin type A and a variable hip abduction orthosis on gross motor function: A randomized controlled trial. *Eur. J. Neurol.* **8**, 109–119 (2001).

55. Gough, M. Continuous postural management and the prevention of deformity in children with cerebral palsy: An appraisal. *Dev. Med. Child Neurol.* **51**, 105–110 (2009).

56. Stott, N. S. & Piedrahita, L. Effects of surgical adductor releases for hip subluxation in cerebral palsy: An AACPDM evidence report. *Dev. Med. Child Neurol.* **46**, 628–645 (2004).

57. Carr, C. & Gage, J. R. The Fate of the Nonoperated Hip in Cerebral Palsy. *J. Pediatr. Orthop.* **7**, 262–267 (1987).

58. Cottalorda, J. *et al.* Predicting the outcome of adductor tenotomy. *Int. Orthop.* **22**, 374–379 (1998).

59. Lameiras-Campagnolo, J. We are IntechOpen , the world ’ s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %. in *Cerebral Palsy - Clinical and Therapeutic Aspects - Chapter 9 - Hip Surgery in Cerebral Palsy* 131–151 (IntechOpen, 2018). doi:http://dx.doi.org/10.5772/intechopen.80442.

60. Al-Ghadir, M., Masquijo, J. J., Guerra, L. A. & Willis, B. Combined femoral and pelvic osteotomies versus femoral osteotomy alone in the treatment of hip dysplasia in children with cerebral palsy. *J. Pediatr. Orthop.* **29**, 779–783 (2009).

61. Dega, W. [Selection of surgical methods in the treatment of congenital dislocation of the hip in children]. *Chir. Narzadow Ruchu Ortop. Pol.* **34**, 357–366 (1969).

62. McNerney, N. P., Mubarak, S. J. & Wenger, D. R. One-stage correction of the dysplastic hip in cerebral palsy with the San Diego acetabuloplasty: results and complications in 104 hips. *J. Pediatr. Orthop.* **20**, 93–103 (2000).

63. Mallet, C. *et al.* One-stage hip reconstruction in children with cerebral palsy: long-term results at skeletal maturity. *J. Child. Orthop.* **8**, 221–228 (2014).

64. Braatz, F. *et al.* Hip-joint congruity after Dega osteotomy in patients with cerebral palsy: long-term results. *Int. Orthop.* **40**, 1663–1668 (2016).

65. Karlen, J. W., Skaggs, D. L., Ramachandran, M. & Kay, R. M. The Dega osteotomy: A versatile osteotomy in the treatment of developmental and neuromuscular hip pathology. *J. Pediatr. Orthop.* **29**, 676–682 (2009).

66. Grudziak, J. S. & Ward, W. T. Dega osteotomy for the treatment of congenital dysplasia of the hip. *J. Bone Jt. Surg. - Ser. A* **83**, 845–854 (2001).

67. Murar, J., Dias, L. S. & Swaroop, V. T. San Diego pelvic osteotomy in patients with closed triradiate cartilage. *J. Child. Orthop.* **12**, 461–471 (2018).

68. Barrie, J. L. & Galasko, C. S. B. Surgery for unstable hips in cerebral palsy. *Journal of Pediatric Orthopaedics Part B* vol. 5 225–231 (1996).

69. Shore, B. J., Powell, D., Miller, P. E., Matheney, T. H. & Snyder, B. D. Acetabular and femoral remodeling after varus derotational osteotomy in cerebral palsy: The effect of age and Gross Motor Function Classification Level. *J. Pediatr. Orthop. Part B* **25**, 322–330 (2016).

70. Root, L., Laplaza, F. ., Brourman, S. . & Angel, D. . The Severely unstable Hip In Cerebral Palsy. *J. Bone Jt. Surg.* **77**, 703–712 (1995).

71. Wynter, M. *et al.* Australian hip surveillance guidelines for children with cerebral palsy: 5-year review. *Dev. Med. Child Neurol.* **57**, 808–820 (2015).

72. Wawrzuta, J. *et al.* Hip health at skeletal maturity : a population-based study of young adults with cerebral palsy. *Dev. Med. Child Neurol.* **58**, 1273–1280 (2016).

73. Henderson, R. C. *et al.* Altered skeletal maturation in moderate to severe cerebral palsy. *Dev. Med. Child Neurol.* **47**, 229–236 (2005).

74. Persson-Bunke, M., Hägglund, G. & Lauge-Pedersen, H. Windswept hip deformity in children with cerebral palsy. *J. Pediatr. Orthop. Part B* **15**, 335–338 (2006).

75. Cliffe, L. *et al.* Correct positioning for hip radiographs allows reliable measurement of hip displacement in cerebral palsy. *Dev. Med. Child Neurol.* **53**, 549–552 (2011).

76. Craven, A., Pym, A. & Boyd, R. N. Reliability of Radiologic Measures of Hip Displacement in a Cohort of Preschool-aged Children With Cerebral Palsy. *J Pediatr Orthop* **34**, 597–602 (2014).

77. Kinch, K. *et al.* How Critical is Patient Positioning in Radiographic Assessment of the Hip in Cerebral Palsy When Measuring Migration Percentage ? *J Pediatr Orthop* **35**, 756–761 (2015).

78. Parrott, J. *et al.* Hip Displacement in Spastic Cerebral Palsy : Repeatability of Radiologic Measurement. *J. Pediatr. Orthop.* **22**, 660–667 (2002).

79. Landis, J. & Koch, G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).

80. Maxim, L. D., Niebo, R. & Utell, M. J. Screening tests: A review with examples. *Inhal. Toxicol.* **26**, 811–828 (2014).

81. Lloyd-Jones, G. Basics of X-ray Physics. *Radiology masterclass* https://www.radiologymasterclass.co.uk/tutorials/physics/x-ray_physics_introduction (2016).

82. Berger, A. Bone mineral density scans. *Bmj* **325**, 484 (2002).

83. Neirynck, J., Proost, R. & Van Campenhout, A. The migration percentage measured on EOS® standing full-leg radiographs: Equivalent and advantageous in ambulant children with cerebral

palsy. *BMC Musculoskelet. Disord.* **20**, 1–6 (2019).

84. Fry, N. R. *et al.* Accurate measurement of muscle belly length in the motion analysis laboratory: Potential for the assessment of contracture. *Gait Posture* **17**, 119–124 (2003).

85. Fry, N. R., Gough, M. & Shortland, A. P. Three-dimensional realisation of muscle morphology and architecture using ultrasound. *Gait Posture* **20**, 177–182 (2004).

86. Barber, L., Barrett, R. & Lichtwark, G. Validation of a freehand 3D ultrasound system for morphological measures of the medial gastrocnemius muscle. *J. Biomech.* **42**, 1313–1319 (2009).

87. Passmore, E., Pandy, M. G., Graham, H. K. & Sangeux, M. Measuring Femoral Torsion In Vivo Using Freehand 3-D Ultrasound Imaging. *Ultrasound Med. Biol.* **42**, 619–623 (2016).

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

89. Committee on Quality Improvement, S. on D. D. of the H. Clinical Practice Guideline : Early Detection of Developmental Dysplasia. *Am. Acad. Pediatr.* **105**, (2000).

90. Smigovec, I., Dapic, T. & Trkulja, V. Ultrasound screening for decentered hips in children with severe cerebral palsy: a preliminary evaluation. *Pediatr. Radiol.* **44**, 1101–1109 (2014).

91. Tegnander, A. & Terjesen, T. Ultrasound measurements in hips of children above 2 years of age: Normal variations in 232 hips. *Acta Orthop. Scand.* **66**, 229–233 (1995).

92. Tegnander, A. & Terjesen, T. Reliability of ultrasonography in the follow up of hip dysplasia in children above 2 years of age. *Acta radiol.* **40**, 619–624 (1999).

93. Thoirs, K. & English, C. Ultrasound measures of muscle thickness: Intra-examiner reliability and influence of body position. *Clin. Physiol. Funct. Imaging* **29**, 440–446 (2009).

94. Oleske, D. M. *Epidemiology and the delivery of health care services: Methods and applications. Chapter 5: Screening and Surveillance for Promoting Population Health*. *Epidemiology and the Delivery of Health Care Services: Methods and Applications* (2009). doi:10.1007/978-1-4419-0164-4.

95. Cuckle, H. Principles of Screening. *Obs. Gynecol.* **6**, (2004).

96. Wilson, J. M. G. & Jungner, G. Principles and Practices of Screening for Disease. *Public Health Pap.* **34**,.

97. Kentish, M., Wynter, M., Snape, N. & Boyd, R. Five-year outcome of state-wide hip surveillance of children and adolescents with cerebral palsy. **4**, 205–217 (2011).

98. Andreo, P. Monte Carlo simulations in radiotherapy dosimetry. *Radiat. Oncol.* **13**, 1–15 (2018).

99. Long, Z. *et al.* Systematic optimization of ultrasound grayscale imaging presets and its application in abdominal scanning. *J. Appl. Clin. Med. Phys.* **21**, 192–199 (2020).

100. Cobelli, C., Carson, E. R., Finkelstein, L. & Leaning, M. S. Validation of simple and complex models in physiology and medicine. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.* **15**, (1984).

101. Reid, S. M., Carlin, J. B. & Reddihough, D. S. Using the Gross Motor Function Classification System to describe patterns of motor severity in cerebral palsy. *Dev. Med. Child Neurol.* **53**, 1007–1012 (2011).

102. Geng, C. *et al.* Using 3-dimensional ultrasound islice technology for the diagnosis of developmental dysplasia of the hip. *J. Ultrasound Med.* **39**, 1117–1123 (2020).

103. Bland, J. M. & Altman, D. G. Validating scales and indexes. *Br. Med. J.* **324**, 606–607 (2002).

104. Hoyt, K., Hester, F. A., Bell, R. L., Lockhart, M. E. & Robbin, M. L. Accuracy of volumetric flow rate measurements: An in vitro study using modern ultrasound scanners. *J. Ultrasound Med.* **28**, 1511–1518 (2009).

105. Chodzyński, K. J. *et al.* An in vitro test bench reproducing coronary blood flow signals. *Biomed. Eng. Online* **14**, 1–26 (2015).

106. Delcker, A., Walker, F., Caress, J., Hunt, C. & Tegeler, C. In vitro measurement of muscle volume with 3-dimensional ultrasound. *Eur. J. Ultrasound* **9**, 185–190 (1999).

107. Obst, S. J., Newsham-West, R. & Barrett, R. S. InVivo measurement of human achilles tendon morphology using freehand 3-D ultrasound. *Ultrasound Med. Biol.* **40**, 62–70 (2014).

108. Robb, J. E. & H??gglund, G. Hip surveillance and management of the displaced hip in cerebral palsy. *J. Child. Orthop.* **7**, 407–413 (2013).

109. Shrader, M. W., Wimberly, L. & Thompson, R. Hip Surveillance in Children With Cerebral Palsy. *J. Am. Acad. Orthop. Surg.* **27**, 760–768 (2019).

110. Porter, D. Patterns of postural deformity in non-ambulant people with cerebral palsy: scoliosis, pelvic obliquity, windswept hips and hip dislocation. *Clin Rehab* 1087–1096 (2007).

111. Hägglund, G., Lauge-Pedersen, H., Persson Bunke, M. & Rodby-Bousquet, E. Windswept hip deformity in children with cerebral palsy: a population-based prospective follow-up. *J. Child. Orthop.* **10**, 275–279 (2016).

112. Culjat, M. O., Goldenberg, D., Tewari, P. & Singh, R. S. A review of tissue substitutes for ultrasound imaging. *Ultrasound Med. Biol.* **36**, 861–873 (2010).

113. Kim, S. M., Sim, E. G., Lim, S. G. & Park, E. S. Reliability of hip migration index in children with cerebral palsy: The classic and modifi ed methods. *Ann. Rehabil. Med.* **36**, 33–38 (2012).

114. Fabry, G., MacEwen, D. & Shands, A. R. Torsion of the femur: a follow-up study in normal and abnormal conditions. *J. Bone Jt. Surg.* **55**-**A**, 1726–1738 (1973).

115. Robin, J. *et al.* Proximal femoral geometry in cerebral palsy. *J. Bone Jt. Surg.* **90**, 1372–1379 (2008).

116. Jozwiak, M., Rychlik, M., Szymczak, W., Grzegorzewski, A. & Musielak, B. Acetabular shape and orintation of the spastic hip in children with cerebral palsy. *Dev. Med. Child Neurol.* **65**, 608–613 (2021).

117. Kotlarsky, P., Haber, R., Bialik, V. & Eidelman, M. Developmental dysplasia of the hip: What has changed in the last 20 years? *World J. Orthop.* **6**, 886–901 (2015).

118. Storer, S. K. & Skaggs, D. L. Developmental dysplasia of the hip. *Am. Fam. Physician* **74**, 1310–1316 (2006).

119. Jin, W. *et al.* Sonographic evaluation of anterosuperior hip labral tears with magnetic resonance arthrographic and surgical correlation. *J. Ultrasound Med.* **31**, 439–447 (2012).

120. Troelsen, A. *et al.* Ultrasound versus Magnetic Resonance Arthrography in Acetabular Labral Tear Diagnostics : A Prospective Comparison in 20 Dysplastic Hips Ultrasound versus Magnetic Resonance Arthrography in Acetabular Labral Tear Diagnostics : A Prospective Comparison in. *Acta radiol.* **1851**, (2009).

121. Sofka, Carolyn, M., Adler, Ronald, S. & Danon, Martha, A. Sonography of the Acetabular Labrum. *J Ultrasound Med* **25**, 1321–1326 (2006).

122. Stone, T. J., Long, N., Petchprapa, C. N. & Adler, R. S. Sonoarthrography of the Hip Labrum: Ultrasound Evaluation of the Anterosuperior Acetabular Labrum Following Joint Distension with Magnetic Resonance Arthrographic Correlation. *J. Orthop. Rheumatol. Sport. Med.* **01**, (2016).

123. Swinscow, T. D. V. 8. The Chi squared tests. *Statistics at Square One* https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests (1997).

124. Middleton, D. “Statistics at square one”. *Br. Med. J.* **2**, 1008–1009 (1976).

125. Cenni, F. *et al.* The reliability and validity of a clinical 3D. *Comput. Methods Programs Biomed.* **136**, 179–187 (2016).

126. Kuhn, T. . *The Structure of Scientific Revolutions*. (University of Chicago Press, 1962).

127. Dobson, F., Boyd, R. N., Parrott, J., Nattrass, G. R. & Graham, H. K. Hip surveillance in children with cerebral palsy. Impact on the surgical management of spastic hip disease. *J. Bone Joint Surg. Br.* **84**, 720–726 (2002).

128. Shore, B. J. *et al.* Adductor Surgery to Prevent Hip Displacement in Children with Cerebral Palsy: The Predictive Role of the Gross Motor Function Classification System. *J. Bone Joint Surg. Am.* 326–334 (2012).

129. Connelly, A., Flett, P., Graham, H. K. & Oates, J. Hip surveillance in Tasmanian children with cerebral palsy. *J. Paediatr. Child Health* **45**, 437–443 (2009).

130. Larkin-Kaiser, K. A. *et al.* Relationship of muscle morphology to hip displacement in cerebral palsy: A pilot study investigating changes intrinsic to the sarcomere. *J. Orthop. Surg. Res.* **14**, 1–9 (2019).

131. Noble, J. J. *et al.* Lower limb muscle volumes in bilateral spastic cerebral palsy. *Brain Dev.* **36**, 294–300 (2014).

132. Vanmechelen, I. M., Shortland, A. P. & Noble, J. J. Lower limb muscle volume estimation from maximum cross-sectional area and muscle length in cerebral palsy and typically developing individuals. *Clin. Biomech.* **51**, 40–44 (2018).

133. Pitcher, C. A. *et al.* Ultrasound characterization of medial gastrocnemius tissue composition in children with spastic cerebral palsy. *Muscle and Nerve* **52**, 397–403 (2015).

134. Multani, I. *et al.* Sarcopenia, cerebral palsy, and botulinum toxin type a. *JBJS Rev.* **7**, 1–10 (2019).

135. Henak, C. R. *et al.* Patient-specific analysis of cartilage and labrum mechanics in human hips with acetabular dysplasia. *Osteoarthr. Cartil.* **22**, 210–217 (2014).

136. Horii, M., Kubo, T., Hachiya, Y., Nishimura, T. & Hirasawa, Y. Development of the acetabulum and the acetabular labrum in the normal child: Analysis with radial-sequence magnetic resonance imaging. *J. Pediatr. Orthop.* **22**, 222–227 (2002).

137. Bsat, S., Frei, H. & Beaulé, P. E. The acetabular labrum: A review of its function. *Bone Jt. J.* **98**-**B**, 730–735 (2016).

138. Noble, J. J. *et al.* Bone strength is related to muscle volume in ambulant individuals with bilateral spastic cerebral palsy. *Bone* **66**, 251–255 (2014).

139. Modlesky, C. M. & Zhang, C. Complicated Muscle-Bone Interactions in Children with Cerebral Palsy. *Curr. Osteoporos. Rep.* **18**, 47–56 (2020).

140. Chang, C. H. *et al.* Dynamic Displacement of the Femoral Head by Hamstring Stretching in Children With Cerebral Palsy. *J Pediatr Orthop* **30**, 475–478 (2010).

9. Appendices

9.1. Appendix 1. Monte Carlo simulation original manuscript

Investigating the Impact of Measurement Uncertainty in the Radiographic Surveillance of Hip Displacement In Cerebral Palsy: A Monte Carlo Simulation

Rebecca Hazell Kay: 1. King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK. 2. One Small Step Gait Laboratory, Guy's & St Thomas' NHS Foundation Trust, London, UK.

Jonathan James Noble: 1. One Small Step Gait Laboratory, Guy's & St Thomas' NHS Foundation Trust, London, UK.

Stephen Keevil: 1. Guy's and St Thomas' NHS Foundation Trust, Medical Physics, London, UK. 2. King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Martin Gough: 1. Evelina London Children's Hospital, Westminster Bridge Road, Lambeth, London, UK

Adam Peter Shortland: 1. King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK. 2. One Small Step Gait Laboratory, Guy's & St Thomas' NHS Foundation Trust, London, UK.

Corresponding author: Rebecca Kay

Address: One Small Step Gait Laboratory, Basement Southwark Wing, Guy's Hospital, Great Maze Pond, London, SE1 9RT.

Telephone: 02071882476.

Fax: NA

Email: Rebecca.east@gstt.nhs.uk.

Running title: Measurement error in hip surveillance

Authors contributions:

Rebecca Kay – lead author, designed and developed the simulation, conducted the analysis and drafted the submission.

Jonathan Noble – helped in the design of the simulation and drafting the submission.

Stephen Keevil – helped in the design of the simulation structure and reviewed and edited the submission.

Martin Gough – helped in defining the decision algorithms in the simulation and reviewed and edited the submission.

Adam Shortland – idea conception, helped in the design of the simulation structure and drafting of the submission.

We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal. We have no conflicts of interest to disclose. All authors have approved the manuscript and agree with its submission.

Abstract

Hip surveillance programmes have greatly improved the management of hip dysplasia in children with cerebral palsy. Reimer's migration percentage is the most common index for quantifying hip dysplasia from planar radiographs. However, measurement uncertainty could undermine the diagnostic accuracy. A Monte Carlo simulation was created to investigate the impact of measurement error on decision making in hip surveillance programmes.

The simulation was designed to mimic the annual surveillance of children with cerebral palsy (Gross Motor Functional Classification System levels III – V) between 2 and 8 years of age. Simulation parameters for the natural history of hip dysplasia and measurement error were derived from published data. At each measurement interval, the influence of uncertainty in the measurement of Reimer's migration percentage on decision-making was investigated.

The probability of a child being indicated for intervention in error during the course of the simulation was relatively high, particularly in the highest functioning cohort where the positive predictive value of Reimer's migration percentage was at best 70% and at worse less than 20%. Including a rate of progression term within the decision-making algorithm had a negative effect on positive predictive power.

This simulation suggests that hip surveillance programmes are sensitive to detecting genuine hip dysplasia but can have poor positive predictive power, potentially resulting in unnecessary indication for intervention.

Introduction

Hip surveillance programmes have been adopted internationally to monitor hip development in children with cerebral palsy[1–4]. Although assessment intervals and measurement variables differ between different programmes, at minimum they comprise a physical examination to assess passive hip abductor range and hip pain, and a radiological assessment. At this assessment an anterior-posterior radiograph is taken, with the individual in a standardised position. The lateral displacement of the femoral head from the acetabulum in the anterior-posterior radiograph is most often estimated using Reimer's migration percentage (RMP)[5]. Simply, the index defines the percentage of the ossified portion of the femoral head that is not covered by the acetabulum at the level of Hilgenreiner's line. The frequency of assessment is often dictated by the level of function of individuals under surveillance, with individuals with Gross motor function classification levels (GMFCS) IV and V receiving annual or in some cases bi-annual assessments, and individuals who can independently mobilise receiving initial assessment and sometimes no further scheduled assessment. The clinical pathway of an individual is defined by the outcome of each assessment. Thresholds for discharge, continued monitoring and referral for orthopaedic management are defined within each programme, typically a threshold for hip displacement and/or progression of hip displacement, a minimum hip abduction range or the presence of hip pain. There is no consensus on the RMP thresholds, but it is widely accepted that hips with RMP of greater than 33% are either at risk or require intervention, and at 50% migration most clinicians would agree that intervention is required. However, measurements of RMP are subject to errors in acquisition and analysis. In the acquisition,the content of an anterior-posteriorX-ray image depends on both the relative orientation of the subject and the X-ray source,and the relative position of the femoral and pelvic segments of the hip[6]. In the analysis, variation in the identification of the required landmarks, differentiation of bony borders and tools used to aid the measurement can result in

both inter and intra assessor variation whichresults in aminimal detectable difference (MDD) of approximately 10% RMP[6–8].

Considering the large MDDs and the relatively low rates of hip displacement (7% RMP across GMFCS levels III to V and as low as 1.3% in the GMFCS level III cohort[9]),it is probable that for some hips the *measured* RMP is significantly different from the actual or *true* RMP, i.e. the hip may be mis-classified as 'at risk' when its positionis satisfactory, or classified as satisfactory when in fact it is 'at risk'.The size of these groups and the impact of mis-classification are under-investigated.

Since the advent of routine monitoring, total dislocation rates have reduced to almost 0%[1]indicating that, when *true* above-threshold RMP is under-estimated at one radiographic assessment, it is likely that at subsequent assessmentsan above threshold measurement will be made. However, there is a potential cohort who are falsely-identified as indicated for intervention, and who, consequently, risk undergoing unnecessary treatment. Members of this group would not be easy to differentiate from the children who had received appropriate intervention and so represent a potential "hidden" population of inappropriately-treated individuals.

In this study we created a Monte Carlo simulation to investigate the influence of uncertainty in the measurement of RMP, specifically, during a prescriptive hip surveillance programme for children with cerebral palsy. We hypothesised that the sensitivity and specificity across the surveillance programme would be high, but that there would be a significant number of cases inappropriately indicated for intervention in a simulated sample population of individuals with cerebral palsy, particularly in those individuals where the underlying rate of progression was low.

Methodology

Description of the simulation

The Monte Carlo simulation described in this paper was developed in Microsoft Excel (Office 365 ProPlus) using Visual Basic for Applications (VBA). It was designed to replicate the radiographic imaging component of a generic hip surveillance programme for non-ambulant individuals with cerebral palsy, and for those who could walk with assistive devices (GMFCS levels III – V) with annual screening between 2 and 8 years of age.

To create representative simulated *true* RMP values (i.e. from hypothetical measurements that were not subject to error), random data points were generated around a normal distribution defined by the mean and standard deviations of RMP values reported at initial assessment by Terjesen[9] for each of the GMFCS levels III to V.Simulated cohorts of 1000 individuals per cohort were created for each of these GMFCS levels. For each *true* RMP value in the simulation, a *measured* RMP value was created by adding a simulated normally-distributed measurement error to the *true* RMP value derived from repeatability data published by Craven *et al*[8]. Craven *et al.* publish the SEM of a single measurement as 3.9%, which corresponds to a MDD of 10.8% This value was chosen as a representative error, similar to others reported in the literature[6,7,10]. Progression of hip displacement was simulated according to a normally-distributed random distribution based on the mean and variance of annual hip progression reported by Terjesen[9]. In this way, we estimated the *true* and *measured* RMP values in a simulated surveillance programme for children with CP (GMFCS III-IV) between the ages of 2 and 8 years, with annual follow-up. Within the simulation, decision making was based on three thresholds, an upper RMP threshold (fixed at 50% RMP throughout), a lower RMP threshold and a progression threshold. Intervention was indicated if the *measured* RMP was greater than a lower threshold and the change in *measured* RMP in successive assessments

exceeded a progression threshold, or the *measured* RMP exceeded the upper (50% RMP) threshold irrespective of progression. A sensitivity analysis was conducted to investigate the impact of varying the lower RMP limit and the progression threshold on the decision to intervene. Those children who were indicated for intervention at any assessment were removed from the simulated programme at that time point.

To assess the stability of the simulation and establish the optimal number of iterations required, the number of simulation repeats was varied, and the results of the simulation recorded. Total number of radiographic assessments conducted during the simulation was chosen as the summary result. This fluctuates depending on whether a positive or negative decision is made, a positive decision results in no further radiographs, whilst a negative result in continuation to the next time point. Stability was defined as the point at which this variable plateaued (variability of less than 2 radiographs) with increasing number of iterations.

		True data points		
		Positive	**Negative**	**Subtotal**
Measured data points	**Positive**	True positive (TP)	False positive (FP)	Intervention indicated group
	Negative	False negative (FN)	True negative (TN)	No intervention indicated group
	Subtotal	Indicated for intervention group	Not indicated for intervention group	Total number of data points

Table 1: Explanation of the possible categorisation of each of the data points when applying the decision algorithm

Data Analysis

To test our hypotheses, the indication for intervention decision was assessed using both the *true* RMP values and the *measured* RMP values at each time point. Table 1 shows the possible

categories of the results. When the simulated *measured* and *true* RMP both satisfied the indications for intervention, the decision was defined as a true positive (the child in the simulation is appropriately indicated for intervention). Similarly, a true negative was defined as an instance where both *measured* and *true* values did not satisfy the indications for intervention (the child is correctly not indicated for intervention). A false positive result occurred when the *measured* data indicated intervention, but the *true* data did not (the child is indicated for intervention when intervention should not be indicated). Similarly, a false negative result was achieved when the *measured* data did not indicate that intervention was necessary, but the *true* data suggested intervention was indicated (a child is not indicated for intervention when intervention should be indicated). From these, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were computed.

To investigate the effect of the lower surgical threshold and progression threshold on the performance of the surveillance programme, simulations were performed at different intervention thresholds for hip displacement and for different rates of hip progression.

Results

Simulation stability

The simulation was stable to within 2 radiographs, from 5000 cycles. All other simulations were run 5000 times, each time with a cohort of 1000 ‘individuals’ and the results averaged.

Hypothesis testing

Table 2 shows the sensitivity, specificity, PPV and NPV of the simulated surveillance programmes. Sensitivity is a measure of the proportion of true positive results that are

correctly identified whilst specificity is the measure of the proportion of true negative results that are correctly identified.

Depending on whether the intervention decision included a progression threshold the sensitivity varied from 0.66 to 0.90. Specificity is very high regardless of the parameters of the intervention decision.

	No progression			*10% progression threshold*		
	GMFCS III	GMFCS IV	GMFCS V	GMFCS III	GMFCS IV	GMFCS V
Sensitivity	0.75	0.82	0.90	0.66	0.80	0.87
Specificity	0.96	0.96	0.96	0.97	0.95	0.95
Postive predictive value (PPV)	0.55	0.78	0.89	0.23	0.63	0.85
Negative predictive value (NPV)	0.96	0.96	0.96	0.97	0.95	0.95

Table 2: Sensitivity, specificity, PPV and NPV across the simulated surveillance programme for each of GMFCS levels III, VI and V. Indication for intervention decision parameters were set at upper intervention limit of 50%, lower limit at 40% and progression threshold at 0% and 10%.

We hypothesised that there would be a significant number of cases that were indicated for intervention as a result of measurement error, and that the proportion of false positives would be greatest in the group with the lowest underlying rate of hip displacement i.e. the GMFCS level III group.

Positive predictive power or value (PPV) is a measure of the probability of a positive result being a true positive result, i.e. aPPV of 20% means that 1 in 5 positive results are truly positive. Figure 1 illustrates the influence of the progression threshold and lower RMP limit on the

positive predictive powerby GMFCS level. Within each GMFCS level, the lower RMP limit does not have a great influence on the PPV. In the GMFCS III cohort, the PPVs vary between 55% and 70% when the progression threshold is set to zero meaning that at this level between 30% and 45% of individuals will be incorrectly indicated for surgery by radiographic imaging. Including a progression threshold in the simulation has a *negative* effect on predictive power particularly in the GMFCS III group.

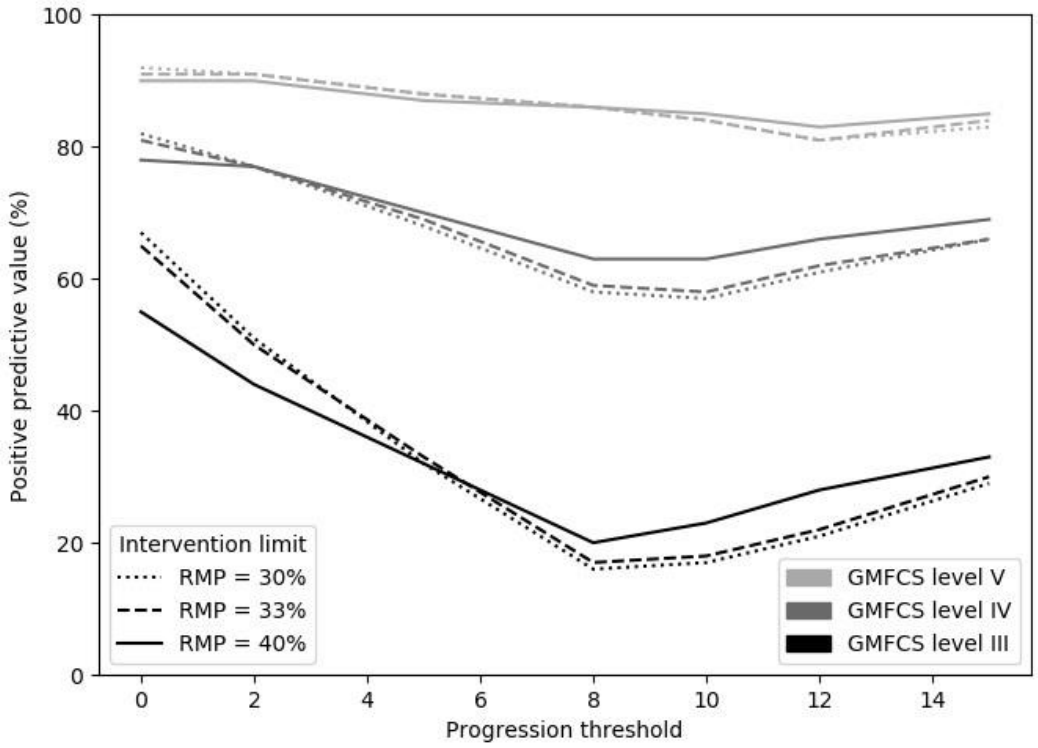


Figure 1: A graph showing the effect of varying progression threshold and lower surgical limit on the predictive power of a positive result.

Discussion

We performed a numerical simulation of the radiographic schedule in a typical hip surveillance programme for children with cerebral palsy to understand the potential influence of

measurement error and decision-making thresholds on the programme's success. Our simulation supported our hypothesis that the surveillance programmes would have high specificity, and that due to the relatively large errors in the measurement of RMP compared to typical hip displacement progression rates, a large number of individuals would be indicated incorrectly for intervention by radiographic measurement (RMP). The proportion of individuals falsely indicated for intervention was particularly high in the GMFCS III group where mean hip progression rates was lower than in GMFCS IV and V groups. However, the sensitivity was, under some conditions, lower than expected.

Simulation validation

Validating a simulation of this nature is a challenge. We cannot hope to model the tacit understanding of the clinicians involved nor all the factors influencing a treatment decision. However, the descriptive validity of the simulation can be assessed by comparing the summary simulation results to the published data that underpins the simulation. Table 3 shows the summary results for the simulation (mean and standard deviation of RMP at initial presentation and annual RMP progression), stratified by GMFCS level compared to the published data[9]upon which the simulation is based.

	Initial Presentation				*Progression*			
	Simulation		Measurements reported in the literature		Simulation		Measurements reported in the literature	
GMFCS level	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
III	26.7	10.8	26.5	10.7	1.3	3.1	1.3	3.1
IV	25.6	20.2	26.2	20.2	3.9	4.9	3.9	4.8
V	29.4	24.2	28.6	24.3	9.5	9.3	9.5	9.4

Table 3: Initial presentation and progression summary results (mean and standard deviation) from the simulation stratified by GMFCS level alongside the published data underpinning the simulation[9]

Predictive validity is a measure of how well the results describedata that were not used to inform the simulation. To assess the predictive validity of the simulation the true positive and false positive results were compared to a published analysis of hip displacement rates. Soo *et al*.[11] published proportions of individuals with hip displacement stratified by GMFCS level. Hip displacement was defined as an RMP of greater than 30%. Our simulation defined indication for intervention thresholds in a similar way. Table 4 shows the rates of indication for intervention from the simulation, comparing these to Soo *et al's* hip displacement rates. For each of the GMFCS levels III-V, the simulation results are similar. When no progression threshold is included, the simulation is slightly less conservative across all GMFCS levels, but when a progression threshold is included in the decision algorithm the simulation becomes more conservative than Soo *et al*'s data.

GMFCS level	Soo *et al* results	Simulation - prog threshold 0%	Simulation - prog threshold 10%
III	43%	45%	26%
IV	69%	71%	62%
V	89%	93%	92%

Table 4: Table showing Soo et al. hip dysplasia rates and simulation indication for intervention rates by GMFCS level. The upper RMP threshold was set at 50% throughout, lower RMP threshold set at 40% and the progression threshold set at 0% and 10% RMP.

Clinical implications

Although, under certain conditions, the sensitivity of hip surveillance programmes appears to be moderate, the design of surveillance programmes means that children with hip displacement who are missed at a single assessment, will most likely be detected at the next assessment without clinically-significant amounts of progression in the interval, thus increasing the detection rate of hip displacement. Due to the large number of true negatives in a surveillance programme, the use of specificity as a measure of programme performance may be flattering. When a positive result does occur, it is important to consider how likely it is that this result is a true positive – this is described by the PPV. High PPV is important when serious interventions are being considered.

To investigate the impact of a false positive result, a further time point was simulated to quantify the number of individuals, falsely indicated for intervention who would progress enough in the course of the following year to pass the threshold for indication for intervention. In this way, it is possible to differentiate those who were simply, prematurely indicated for intervention and those who were falsely indicated for intervention and would still not have been indicated for intervention at the subsequent time point. Figure 2 shows the premature indication for intervention rate within the false positive group by GMFCS level. In the GMFCS V cohort we can see that 60%-80% of those falsely classified as indicated for intervention were merely pre-emptive. However, in the GMFCS III group only 10%-35% of the total false positive group were pre-emptive, indicating that the majority of those falsely indicated for intervention were not indicated for intervention at the next time point. Where the rates of hip displacement are slower, applying a progression threshold increases the chance of misclassification (figure 1). Measuring progression requires comparing radiographs from different time points, typically one year apart. Therefore, there are two instances where

measurement uncertainties are introduced. In the GMFCS III cohort, when a positive result is indicated it should be treated with caution, particularly if progression is considered in the clinical decision-making.

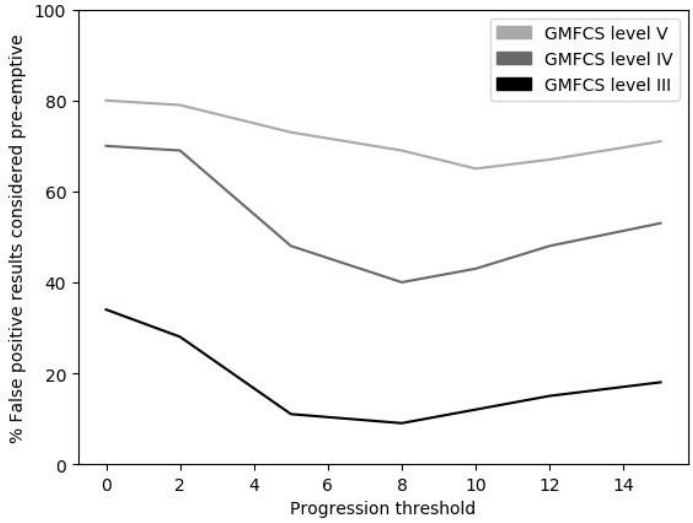


Figure 2: A graph showing how "pre-emptive" indication for intervention rates vary with Gross Motor Function Classification System level and progression threshold

The American academy of Cerebral Palsy and Developmental Medicine (AACPDM) guidelines[4] suggest bi-annual follow up for most of this cohort until the age of 5. However, increasing the frequency of assessment will increase the chance of mis-classification. On average the progression of hip displacement in individuals with cerebral palsy is below 10% RMP per year, which is comparable in magnitude to the measurement's MDD. Sampling more frequently will minimise the amount of true progression between monitoring points, increasing the chance of measurement error influencing the decision. It is important to ensure that monitoring intervals are optimised with regard to the expected progression rates of individuals to limit misclassification rates, whilst ensuring that individuals with high progression rates are detected in a timely manner.

Whilst hip surveillance programmes are not screening programmes, some of the principles of screening programmes can be applied. In both screening and surveillance, it is important to ensure that no individual who truly has the condition is missed. A sensitive, yet economic and simple to administer test is adopted. In a screening program a secondary highly-specific test is then applied to confirm a positive result. Once there is a positive radiographic result, particularly in the GMFCS level III cohort, it may be advisable to seek further imaging which better captures the positions of the acetabulum and femoral head.

Limitations

This is a mathematical model of a clinical scenario, and therefore has limitations. Perhaps the most significant is the assumption that the underpinning data is normally distributed (see for example Terjesen[9]). Secondly, Hermanson *et al*[12] showed that age and RMP at initial presentation are risk factors for progression of hip displacement. This was not included in our simulation.

Error data were assumed to be normally distributed with zero systematic bias, however we know that a systematic error due to X-ray absorption in different tissues is likely to exist. Bone absorbs X-rays much more than the surrounding tissues, resulting in high contrast images of the skeleton in the adult. However, in the infant, the bones of the hip are largely cartilaginous, with the development of ossified bone occurring as the child matures[13]. Systematic measurement errors in planar radiographic imaging are likely as the contrast between non-ossified bone and other tissues is less clear. Unfortunately, we do not know enough about the development of ossification in the hips of children with CP to quantify this

error and its potential influence RMP measurements. There is a potential variation in reliability of RMP measurements with age[7], however Craven et al did not find a significant difference in the repeatability of RMP between their age bands[8]. An age-dependent error function was not

included in our model due to insufficient published data. The SEM of measurement used in this simulation is inline with reported values in the literature which are based on measurements taken across a full age spectrum of children.

Summary

This paper indicates that there is a population of children who may be indicated for interventions for hip displacement due to measurement error from planar radiograph. The size of the group is influenced the underlying rate of hip displacement and parameters used to define a treatment decision (critical levels of hip displacement and progression). There is an increased risk of misclassification when measurement from radiographs at successive time points are compared, particularly when the underlying rate of hip displacement is low. Indications for intervention from planar radiographs, in individuals categorised as GMFCS III, should be treated with caution and further investigations should be considered.

Acknowledgements

Thanks to Dr. Robert Eckersley, Dr. Anne McNee and Professor Lionel Tarassenko for their guidance and advice on the construction of this simulation. This work was funded by Action Medical Research (UK), registered charity in England & Wales (208701) and in Scotland (SC039284), grant number GN2532. This work was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z].

References

1. Hagglund, G., Alriksson-Schmidt, A., Rodby-Bousquet, E., Wagner, P. & Westbom, L. Prevention of dislocation of the hip in children with cerebral palsy: 20 year results

of a population based prevention programme. *Bone Jt. J.* **96**, 1546–1552 (2014).

2. Kentish, M., Wynter, M., Snape, N. & Boyd, R. Five-year outcome of state-wide hip surveillance of children and adolescents with cerebral palsy. **4**, 205–217 (2011).

3. NHS Scotland. Cerebral Palsy Integrated Pathway Scotland (CPIPS). (2017). Available at: https://apcp.csp.org.uk/system/files/cpips_manual_2017.pdf. (Accessed: 1st March 2019)

4. O'Donnell, M. *et al. Hip surveillance - Bottom Line 'Evidence-Informed' Recommendations for the Hip Surveillance in Individuals with Cerebral Palsy*. (2016).

5. Reimers, J. *The stability of the hip in children. A radiological study of the results of muscle surgery in cerebral palsy. Acta orthopaedica Scandinavica. Supplementum* **184**, (1980).

6. Cliffe, L. *et al.* Correct positioning for hip radiographs allows reliable measurement of hip displacement in cerebral palsy. *Dev. Med. Child Neurol.* **53**, 549–552 (2011).

7. Parrott, J. *et al.* Hip Displacement in Spastic Cerebral Palsy : Repeatability of Radiologic Measurement. *J. Pediatr. Orthop.* **22**, 660–667 (2002).

8. Craven, A., Pym, A. & Boyd, R. N. Reliability of Radiologic Measures of Hip Displacement in a Cohort of Preschool-aged Children With Cerebral Palsy. *J Pediatr Orthop* **34**, 597–602 (2014).

9. Terjesen, T. The natural history of hip development in cerebral palsy. *Dev. Med. Child Neurol.* **54**, 951–957 (2012).

10. Kinch, K. *et al.* How Critical is Patient Positioning in Radiographic Assessment of the Hip in Cerebral Palsy When Measuring Migration Percentage ? *J Pediatr Orthop* **35**, 756–761 (2015).

11. Soo, B. *et al.* Hip Displacement in Cerebral Palsy. *J. Bone Jt. Surg.* **88**, 121 (2006).

12. Hermanson, M., Hägglund, G., Riad, J., Rodby-Bousquet, E. & Wagner, P. Prediction of hip displacement in children with cerebral palsy DEVELOPMENT OF THE CPUP HIP SCORE. **97**, 1441–1444 (2015).

13. Osborne, D., Effmann, E., Broda, K. & Harrelson, J. The development of the upper end of the femur, with special reference to its internal architecture. *Radiology* **137**, 71–76 (1980).

9.2. Appendix 2. Regression output from rotational sensitivity experiments

Experimental set up	R2	Regression output

FHPPR_R — 0.94

Flex/Ext

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.045	.003		16.553	.000	.039	.051
	Degrees	.001	.000	.970	12.571	.000	.001	.002

a. Dependent Variable: Felx_Ext

FHPPR_R — 0.005

Ab/Adduction

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-.010	.003		-3.761	.002	-.016	-.005
	Degrees	-3.275E-5	.000	-.071	-.283	.781	.000	.000

a. Dependent Variable: Ab_Add

FHPPR_R — 0.119

Int/Ext rotation

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.001	.017		.040	.968	-.035	.036
	Degrees	.001	.001	.345	1.471	.161	.000	.003

a. Dependent Variable: Int_Ext

FHPPR_R — 0.520

Int/Ext rotation (outlier removed)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.017	.002		9.946	.000	.014	.021
	Degrees	.000	.000	.721	4.032	.001	.000	.000

a. Dependent Variable: Int_Ext

FHPPR_L — 0.92

Flex/Ext

(+40° ext rotation)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.120	.010		12.227	.001	.089	.152
	Degrees	.002	.000	.961	6.003	.009	.001	.004

a. Dependent Variable: Flex_Ext

FHPPR_L — 0.701

Ab/Adduction

(+40° ext rotation)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.140	.004		35.037	.000	.130	.149
	Degrees	-.001	.000	-.837	-4.047	.005	-.001	.000

a. Dependent Variable: Ab_Add

FHPPR_L 0.633

Int/Ext rotation

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.181	.006		31.418	.000	.167	.194
	Degrees	.001	.000	.795	3.472	.010	.000	.001

a. Dependent Variable: Int_Ext

LHC_R 0.647

Flex/Ext

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.560	.008		67.421	.000	.541	.578
	Degrees	-.002	.000	-.804	-4.486	.001	-.003	-.001

a. Dependent Variable: Flex_Ext

LHC_R 0.871

Ab/Adduction

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.614	.008		72.477	.000	.596	.633
	Degrees	.005	.001	.933	8.232	.000	.003	.006

a. Dependent Variable: Ab_Add

LHC_R 0.787

Int/Ext rotation

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.562	.004		153.453	.000	.554	.570
	Degrees	-.001	.000	-.887	-7.445	.000	-.001	-.001

a. Dependent Variable: Int_Ext

LHC_L 0.882

Flex/Ext

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.787	.006		137.843	.000	.771	.803
	Degrees	-.001	.000	-.939	-5.458	.005	-.002	-.001

a. Dependent Variable: Flex_Ext

LHC_L 0.708

Ab/Adduction

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.789	.007		119.218	.000	.775	.804
	Degrees	.002	.000	.841	5.826	.000	.001	.003

a. Dependent Variable: Ab_Add

LHC_L 0.953

Int/Ext rotation

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.802	.004		210.767	.000	.793	.811
	Degrees	-.002	.000	-.976	-12.767	.000	-.002	-.002

a. Dependent Variable: Int_Ext

LHC_L Flex/Ext (+40° ext rotation) 0.579

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.844	.007		116.446	.000	.826	.863
	Degrees	-.001	.000	-.761	-2.624	.047	-.002	.000

a. Dependent Variable: Flex_Ext

LHC_L Ab/Adduction (+40° ext rotation) 0.926

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.866	.008		104.286	.000	.845	.886
	Degrees	.005	.001	.962	8.679	.000	.004	.007

a. Dependent Variable: Ab_Add

9.3. Appendix 3. Regression output from translational sensitivity experiments

Experimental set up	R2	Regression output

Right side lying LHC ML displacement 0.99

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.676	.005		140.002	.000	.666	.685
	ML	-.027	.000	-.995	-56.563	.000	-.028	-.026

a. Dependent Variable: LHC

Left side lying LHC ML displacement 0.99

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.416	.004		95.003	.000	.407	.425
	ML	-.028	.000	-.996	-63.304	.000	-.029	-.027

a. Dependent Variable: LHC

Right side lying LHC AP displacement 0.89

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.517	.008		63.128	.000	.499	.534
	AP	-.025	.002	-.946	-10.486	.000	-.031	-.020

a. Dependent Variable: LHC

Left side lying LHC AP displacement 0.92

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	.381	.013		29.512	.000	.354	.408
	AP	-.028	.002	-.960	-16.463	.000	-.032	-.025

a. Dependent Variable: LHC

Right supine FHPPR ML displacement 0.39

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.013	.018		.718	.482	-.025	.050
	ML	.011	.003	.625	3.396	.003	.004	.018

a. Dependent Variable: FHPPR

Left supine FHPPR ML displacement 0.66

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-.014	.002		-5.883	.000	-.019	-.009
	ML	.004	.001	.814	5.057	.000	.002	.005

a. Dependent Variable: FHPPR

Right supine FHPPR AP displacement 0.99

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-.010	.003		-3.853	.001	-.016	-.005
	AP	.024	.000	.997	67.676	.000	.023	.025

a. Dependent Variable: FHPPR

Left supine FHPPR AP displacement 0.98

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-.021	.005		-4.510	.000	-.031	-.011
	AP	.025	.001	.991	34.682	.000	.023	.026

a. Dependent Variable: FHPPR

9.4. Appendix 4. Ethics application approval and supporting documentation

9.4.1. REC favourable opinion letter

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Gwasanaeth Moeseg Ymchwil
Research Ethics Service

Ariennir gan Lywodraeth Cymru
Funded by Welsh Government

WALES REC 7
PO Box 108
Building 1
St David's Park
Jobswell Road
Carmarthen
SA31 3WY

Tel: 01267 225045
Email: sue.byng@wales.nhs.uk

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

Dr Adam Shortland
One Small Step Gait Laboratory
Guy's Hospital, Basement Southwark Wing
Great Maze Pond
London
SE19RT

5 April 2017

Dear Dr Shortland

Study title:	**Monitoring Hip Dysplasia in Cerebral Palsy with Three-Dimensional Ultrasound.**
REC reference:	**17/WA/0093**
IRAS project ID:	**220978**

The Proportionate Review Sub-committee of the Wales REC 7 reviewed the above application on 05 April 2017.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact hra.studyregistration@nhs.net outlining the reasons for your request. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

1) Clarification is requested whether the age appropriate information sheets take into consideration potential learning difficulties or whether a separate information sheet is required.
2) The Parent/Guardian information sheet should state the study is being undertaken to fulfil an educational qualification.
3) The information sheets should reassure the participants that having an ultrasound scan is not painful.

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Summary of discussion at the meeting (if applicable)

Recruitment arrangements and access to health information, and fair participant selection

The PRSC asked how the 12 people would be recruited for the repeatability study.
The Chief Investigator responded they had some difficulty in recruiting children who had to attend for repeated things for the research. Therefore, the first 12 recruited into the main study who agreed to be researched on would be the same 12 children who would be in the repeatability study.
The PRSC noted this was included in the PIS.

Care and protection of research participants; respect for potential and enrolled participants' welfare and dignity

The PRSC discussed whether the x-rays would be part of normal clinical care or whether they should be regarded as a component of the research which therefore required consent. It was concluded that the x-rays were standard practice and that no more ionising radiation would be involved and therefore there was no requirement to outline the use of x-rays within the research.

Informed consent process and the adequacy and completeness of participant information

The PRSC commented that some children, though by no means all, had learning difficulties associated with their cerebral palsy and noted the researchers had provided chronological age group information sheets but wondered if possible learning difficulties had been factored into the provision of the information sheets.

The PRSC noted the Parent/Guardian PIS did not mention that the study was being undertaken to fulfil an educational qualification. It was not felt necessary to include this information in the PIS for the children.

The PRSC felt that the information sheets should make it clear that having an ultrasound scan was not painful.

Suitability of the applicant and supporting staff

The PRSC commented it was unclear whether the student researcher would be involved in taking and/or interpreting the ultrasound or only comparing the various indices.
It was confirmed the student researcher was a clinical scientist who had the capability to carry out these tasks.
The PRSC was satisfied with this explanation.

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
IRAS Application Form [IRAS_Form_22032017]		22 March 2017
Letter from funder [Research Agreement]		29 November 2016
Letters of invitation to participant	1.1	31 January 2017
Participant consent form [Assent Form]	1.2	Undated
Participant consent form [Consent To Contact]	1.2	16 March 2017
Participant consent form [Parents]	1.3	16 March 2017
Participant information sheet (PIS) [Parent Validation]	1.4	16 March 2017
Participant information sheet (PIS) [Parent Repeatability]	1.4	16 March 2017
Participant information sheet (PIS) [Child 2-4Years]	1.2	16 March 2017
Participant information sheet (PIS) [Child 5-7 Repeatability]	1.2	16 March 2017
Participant information sheet (PIS) [Child 5-7 Validation]	1.2	16 March 2017
Participant information sheet (PIS) [Child 8-11 Repeatability]	1.2	16 March 2017
Participant information sheet (PIS) [Child 8-11 Validation]	1.2	16 March 2017
Participant information sheet (PIS) [Child 12-16 Repeatability]	1.2	16 March 2017
Participant information sheet (PIS) [Child12-16 Validation]	1.2	16 March 2017
Referee's report or other scientific critique report [Referee A]	1.0	
Referee's report or other scientific critique report [Referee B]	1.0	
Referee's report or other scientific critique report [Referee C]	1.0	
Referee's report or other scientific critique report [Referee D]	1.0	
Research protocol or project proposal	1.2	13 March 2017
Summary CV for Chief Investigator (CI) Dr Adam Shortland		16 March 2017
Summary CV for student [Rebecca East]		

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

17/WA/0093	**Please quote this number on all correspondence**

Yours sincerely

pp. Dr John Buchan
Vice-Chair

Enclosures: *List of names and professions of members who took part in the review*

"After ethical review – guidance for researchers"

Copy to: *Ms Elizabeth Bruna*
Dr Mays Jawad, Guy's and St Thomas' Foundation NHS Trust

Wales REC 7

Attendance at PRS Sub-Committee of the REC meeting on 05 April 2017

Committee Members:

Name	*Profession*	*Present*	*Notes*
Dr John Buchan	Retired Medical Practitioner /Vice-Chair	Yes	
Dr Raymond Jones	Lay member	Yes	
Mrs Rosemary Whittemore	Lay member	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Ms Sue Byng	REC Manager

9.4.2. HRA approval letter

NHS
Health Research Authority

Dr Adam Shortland
One Small Step Gait Laboratory, Guy's Hospital, Basement
Southwark Wing, Great Maze Pond,
London
SE1 9RT

Email: hra.approval@nhs.net

20 April 2017

Dear Dr Shortland

Letter of HRA Approval

Study title:	**Monitoring Hip Dysplasia in Cerebral Palsy with Three-Dimensional Ultrasound.**
IRAS project ID:	**220978**
REC reference:	**17/WA/0093**
Sponsor	**Guy's and St Thomas' Foundation NHS Trust**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details

and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators"*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/.

IRAS project ID	220978

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is **220978**. Please quote this on all correspondence.

Yours sincerely

Miss Helen Penistone
Assessor

Email: hra.approval@nhs.net

Copy to: *Ms Elizabeth Bruna*
Dr Mays Jawad, Guy's and St Thomas' Foundation NHS Trust

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	*Version*	*Date*
IRAS Application Form [IRAS_Form_22032017]		22 March 2017
Letter from funder [ResearchAgreement]	1.0	29 November 2016
Letters of invitation to participant [ParticpantInvitation]	1.1	31 January 2017
Participant consent form [AssentForm]	1.2	-
Participant consent form [ConsentToContact]	1.2	16 March 2017
Participant consent form [ParentsConsent]	1.4	10 April 2017
Participant information sheet (PIS) [PISChild2-4Years]	1.2	16 March 2017
Participant information sheet (PIS) [PIS Parent validation study]	1.5	10 April 2017
Participant information sheet (PIS) [PIS Parent repeatability study]	1.5	10 April 2017
Participant information sheet (PIS) [PISChild5-7Repeatability]	1.3	April 2017
Participant information sheet (PIS) [PISChild5-7Validation]	1.3	April 2017
Participant information sheet (PIS) [PISChild8-11Repeatability]	1.3	April 2017
Participant information sheet (PIS) [PISChild8-11Validation]	1.3	April 2017
Participant information sheet (PIS) [PISChild12-16Repeatability]	1.3	April 2017
Participant information sheet (PIS) [PISChild12-16Validation]	1.3	April 2017
Referee's report or other scientific critique report [RefereeA]	1.0	
Referee's report or other scientific critique report [RefereeB]	1.0	
Referee's report or other scientific critique report [RefereeC]	1.0	
Referee's report or other scientific critique report [RefereeD]	1.0	
Research protocol or project proposal [Hip Dysplasia Validation and Repeatability Protocol]	1.2	13 March 2017
Summary CV for Chief Investigator (CI) [CV_Shortland_ChiefInvestigator]		16 March 2017
Summary CV for student [CV Rebecca East]		

Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the, *participating NHS organisations, capacity and capability* and *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: Dr Adam Shortland
Tel: 02071882476
Email: adam.shortland@gstt.nhs.uk

HRA assessment criteria

Section	HRA Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	All exposures to ionising radiation are according to clinical need and not influenced by the research protocol.
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	This is a single site study taking place in the NHS where the site is also the study sponsor. Therefore, no agreement is expected.
4.2	Insurance/indemnity arrangements assessed	Yes	Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this

Section	HRA Assessment Criteria	Compliant with Standards	Comments
			research study
4.3	Financial arrangements assessed	Yes	Funding to support the study has been granted by Action Medical Research.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	No comments
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.
This is a single site study. All research activities will take place at site as per the protocol. If this study is subsequently extended to other NHS organisation(s) in England, an amendment should be submitted to the HRA, with a Statement of Activities and Schedule of Events for the newly participating NHS organisation(s) in England. The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local

LCRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.
This is a single site study sponsored by the site. The R&D office will confirm to the CI when the study can start.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).
The Chief Investigator will act as Principal Investigator at site to oversee the research activities. GCP training is not a generic training expectation, in line with the HRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
Where arrangements are not already in place: Research team members undertaking activities in A19 that would have a direct bearing on the quality of care would be expected to obtain an honorary research contract from the NHS organisation. This would be on the basis of a Research Passport (if university employed) or an NHS to NHS confirmation of pre-engagement checks letter (if NHS employed). These should confirm enhanced DBS checks, including appropriate barred list checks (if activity is undertaken regularly with children), and occupational health clearance. For research team members taking consent or conducting an ultrasound assessment which would have no direct bearing on the quality of care, a Letter of Access based on enhanced DBS checks, including appropriate barred list checks (if activity is undertaken regularly with children), and occupational health clearance would be appropriate.

IRAS project ID	220978

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.
The applicant has indicated that they intend to apply for inclusion on the NIHR CRN Portfolio.

9.4.3. Letter of invitation

Guy's and St Thomas' NHS
NHS Foundation Trust

EvelinaLondon

Guys' Hospital
Great Maze Pond
London SE1 9RT
Hospital No: 020 7188 7188
Direct Line: 020 7188 2476
Email: adam.shortland@gstt.nhs.net

DATE

Dear

I am writing to you to ask if you and your child would be interested in participating in a study we are conducting at the One Small Step Gait Laboratory, Guy's Hospital.

This study aims to validate the use of 3D ultrasound system to image the hip in children with cerebral palsy. Currently we use x-rays to monitor hip development in these children.

To validate the system, we need volunteers with cerebral palsy who have undergone hip x-rays as part of their routine clinical care. We will need to take ultrasound images of their hips within a month of their hip x-ray so that we can compare the images from the two imaging techniques.

If you and your child are interested in this study, please find some more detailed information sheets enclosed.

Please contact us on the numbers or email address above if you have any questions or would like to get involved in our study.

Kind Regards,

Dr Adam Shortland

Consultant Clinical Scientist

9.4.4. Patient information sheet (example of parent and child aged 5-7 years)

Guy's and St Thomas' NHS
NHS Foundation Trust

EvelinaLondon

Guys' Hospital
Great Maze Pond
London SE1 9RT
Hospital No: 020 7188 7188
Direct Line: 020 7188 2476
Email: adam.shortland@gstt.nhs.net

PARENT/GUARDIAN INFORMATION SHEET - Validation

Study Title: Monitoring hip dysplasia in children with Cerebral Palsy.

Introduction
Your child is invited to take part in a research study. Before you decide whether you want your child to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends and relatives. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you and your child wish to take part.

What is the purpose of this study?
In this study, we hope to develop a system based on 3D ultrasound to assess the hip dysplasia (mal-development of the hip). Currently, we use x-rays to monitor patients who are at risk. However routine use of x-rays exposes children to radiation which can be harmful. Diagnostic ultrasound is commonly used to scan pregnant women to look at their unborn child and there are no known side effects. The system we hope to develop will also allow us to study the hip in 3D (x-ray images are 2D). We think by studying in greater detail the hips of patients who are at risk of hip dysplasia, we may be able to better identify those at risk of hip dislocation. We will compare the x-ray and 3D ultrasound scans to determine if 3D ultrasound can replace x-ray images as a routine monitoring tool. We may also ask if you and your child would like to be involved in another study we are conducting which is looking at the repeatability of the 3D ultrasound assessment.

Why has your child been chosen?
You have been sent this information as your child has or is undergoing a routine clinical X-ray to assess their hip development. We are asking 40 individuals aged between 2-16 years old who have had or are having a routine clinical X-ray within a month to participate in this study.
If you are interested in your child participating in our study, then we would appreciate you letting us know so that we can arrange to talk further about the study or a convenient time for you and your child to attend for an appointment.

What we want to do
We want to test a 3D ultrasound system that we have developed. We intend to do this by taking images of your child's hips and comparing these images to the X-ray images collect as part of their routine clinical management. This will involve your child attending for a 30-minute appointment within 1 month of their routine clinical X-ray. This appointment will be arranged at a mutually convenient time and conducted at Guy's and St Thomas' Hospital either on the Guy's site or the Evelina site.

What is being asked of you?

We would ask you to bring your child to an appointment to undergo the 3D ultrasound assessment. This will involve your child lying in a few different positions on the assessment couch and having several scans of their hip conducted. The ultrasound is a totally safe imaging technique with no side effects of its use. The positioning of your child will also be guided by their comfort. We will also ask for permission to access your child's clinically acquired hip X-ray for comparison purposes.

3D Ultrasound
We will need access to your child's hip/s. We will ask your child to wear some shorts, we expect that we will access the hip by pulling the short leg up. We will ask your child to lie on the assessment couch in several different positions which will be determined by their comfort. We will scan around the hip and surrounding musculature. We may need to repeat scans and take some scans from different angles. We anticipate that the scanning will take up to 20 minutes. Your child will need to lie relatively still during a scan but will have the opportunity to move about between scans. We anticipate that each scan will last no more than 2 minutes.

What are the possible risks and benefits?
Ultrasound is considered a safe imaging technique. Ultrasound is used during pregnancy to image the foetus. There are no direct benefits from your child taking part in this study.

Who decides if my child will take part?
The decision to take part is up to you and your child. As long as your child has cerebral palsy and has undergone a recent (within the last month) hip x-ray and has not had surgery to both hips they are eligible to take part in this study. If you decide to take part, you will be asked to sign a consent form.

Is this study being conducted as part of an educational project?
Yes this project will contribute to a doctoral level qualification.

Can you change your mind?
Yes, if you decide that you do not want to participate – that's OK you can withdraw your child from the study at any point without giving a reason. Any data collected up until that point may still be used unless you request for it not to be. Your child's medical care will not be affected in anyway.

What about confidentiality?
All information collected during the course of this study will be kept strictly confidential. Any information leaving the hospital will have your personal details removed so that your child cannot be identified from it.
The research team will have access to the data collected as part of this study. All data will be stored on a Trust network. On closure of the study your child's data will be kept until they are 25 years old in accordance with Trust policy.

Are there any costs?
Yes. You may incur some travelling expenses for which you will be reimbursed.

Who has reviewed this study?
This study has been reviewed by the National Research Ethics Service at XX.

What if something goes wrong?
If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. Adam Shortland is the Principal Investigator (02071882476, adam.shortland@gstt.nhs.uk). If you remain unhappy and wish to complain formally, you can do this through the Guy's and St Thomas' Patients Advice and Liaison Service (PALS) on 020

7188 8801, pals@gstt.nhs.uk. The PALS team are based in the main entrance on the ground floor at St Thomas' Hospital and on the ground floor at Guy's Hospital in the Tower Wing.

In the event that something does go wrong and you are harmed during the research and this is due to someone's negligence then you may have grounds for legal action for compensation against Guy's and St Thomas' Foundation NHS Trust but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you (if appropriate).

What should be done with this information sheet?
Please take your time to read this information carefully. Keep it for reference if you and your child decide to take part.

When you have made a decision
If you decide to participate, please contact Becky East on 020871882476 or Rebecca.east@gstt.nhs.uk and she will organise an appointment for the 3D ultrasound appointment and a mutually convenient time. You will be asked to sign a consent form at the appointment.

Have you got any questions?
If you would like any more information about the study or have any concerns, please contact:

Adam Shortland, Consultant Clinical Scientist on 020 7188 2476 or adam.shortland@gstt.nhs.uk.
Becky East, Clinical Scientist on 0208 7188 2476 or Rebecca.east@gstt.nhs.uk.

Guy's and St Thomas' NHS
NHS Foundation Trust

Evelina London

Guys' Hospital
Great Maze Pond
London SE1 9RT
Hospital No: 020 7188 7188
Direct Line: 020 7188 2476
Email: adam.shortland@gstt.nhs.net

CHILD (5-7 YEARS) INFORMATION SHEET

Study Title: Monitoring hip dysplasia in children with Cerebral Palsy.

Please will you help us with our work?

We would like to take a picture of your hip with a machine called an ultrasound scanner. It will not hurt.

Please ask us or your mummy/daddy or carer questions.

9.4.5. Consent form

Guy's and St Thomas' NHS
NHS Foundation Trust

Evelina London
Children's Hospital

Guys' Hospital
Great Maze Pond
London
SE1 9RT
Hospital No: 020 7188 7188
Direct Line: 020 7188 2476
Email: adam.shortland@gstt.nhs.net

PARENT/GUARDIAN CONSENT FORM

This study received favourable opinion from the WALES REC 7 Ethics committee.

Study Number: REC Number 17/WA/0093, IRAS number 220978
Study Title: **Monitoring hip dysplasia in children with Cerebral Palsy.**
Researchers: Adam Shortland, Rebecca East, Jonathan Noble, Martin Gough

Please initial box

1. I confirm that I have read and understand the parent/guardian information sheet dated April 2017 (v.1.3) for the above study and my child and I have had the opportunity to ask questions. ☐
2. I understand that my child's participation is voluntary and that I am free to withdraw my child at any time, without giving any reason and without my child's medical care or legal rights being affected. ☐
3. I understand that relevant sections of my child's medical notes and data collected during the study, may be looked at by individuals from the research team, where it is relevant to my child taking part in this research. I give permission for these individuals to have access to my child's records. ☐
4. I agree to allow my child take part in the above study. ☐
5. I agree to allow the data collected from my child to be used in other relevant ethically approved studies in the future. ☐
6. I agree that images and data collected during this study may be published. All data will be anonymised prior to publication. *(optional)* ☐

_____
Name of Child

_____	_____	_____
Name of Parent/Guardian	Date	Signature
_____	_____	_____
Name of Person taking consent	Date	Signature

This form will be kept in the research file, a copy of this consent will be given to you for your personal records

9.5. Appendix 5. Validation of LHC – original article

DEVELOPMENTAL MEDICINE & CHILD NEUROLOGY ORIGINAL ARTICLE

3D ultrasound to quantify lateral hip displacement in children with cerebral palsy: a validation study

REBECCA H KAY[1,2] | JONATHAN J NOBLE[2] | LIAM JOHNSTON[1,3] | STEPHEN F KEEVIL[1,4] | MICHAIL KOKKINAKIS[5,6] | DANIEL REED[5] | MARTIN GOUGH[5] | ADAM P SHORTLAND[1,2]

1 Biomedical Engineering and Imaging Sciences, King's College London, London; **2** One Small Step Gait Laboratory, Guy's & St Thomas' NHS Foundation Trust, London; **3** King's College Hospital, London; **4** Medical Physics, Guy's and St Thomas' NHS Foundation Trust, London; **5** Evelina London Children's Hospital, London; **6** Faculty of Life Sciences and Medicine, King's College London, London, UK.

Correspondence to Rebecca H Kay at One Small Step Gait Laboratory, Basement Southwark Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK. E-mail: Rebecca.east@kcl.ac.uk

PUBLICATION DATA
Accepted for publication 1st July 2020.
Published online

ABBREVIATIONS

LHC	Lateral head coverage
MDD	Minimal detectable difference
RMP	Reimer's migration percentage

AIM To assess the validity of a new index, lateral head coverage (LHC), for describing hip dysplasia in a population of children with cerebral palsy (CP).
METHOD LHC is derived from 3D ultrasound assessment. Twenty-two children (15 males, seven females; age 4–15y) with CP undergoing routine hip surveillance were recruited prospectively for the study. Each participant had both a planar radiograph acquired as part of their routine care and a 3D ultrasound assessment within 2 months. Reimer's migration percentage (RMP) and LHC were measured by the same assessor, and the correlation between them calculated using Pearson's correlation coefficient. The repeatability of LHC was investigated with three assessors, analysing each of 10 images three times. Inter- and intra-assessor variation was quantified using intraclass correlation coefficients.
RESULTS LHC was strongly correlated with RMP (Spearman's rank correlation coefficient=–0.86, $p<0.001$). LHC had similar inter-assessor reliability to that reported for RMP (intraclass correlation coefficient=0.97 and intra-assessor intraclass correlation coefficient=0.98).
INTERPRETATION This is an initial validation of the use of 3D ultrasound in monitoring hip development in children with CP. LHC is comparable with RMP in estimating hip dysplasia with similar levels of reliability that are reported for RMP.

Hip dysplasia is a common developmental problem for children with cerebral palsy (CP) with a prevalence of around 35% to 40%.[1,2] Hip dysplasia has the potential to cause severe pain and reduced function and, in extreme cases, can progress to complete dislocation.[3,4] The risk factors for hip dysplasia and dislocation include age,[5–8] subtype of CP,[7,8] and motor function, with patients in Gross Motor Function Classification System (GMFCS) level V being at greatest risk.[2,5,8] Progression to symptomatic hip dysplasia and dislocation can be prevented using a variety of surgical interventions.

To reduce the risk of significant hip dysplasia, surveillance programmes have been developed to monitor children with CP. These programmes rely on measurements made from planar radiographs of the hip to quantify hip displacement and acetabular development. Different indices have been developed, however Reimer's migration percentage[9] (RMP) is the most widely adopted. RMP is measured from a planar radiograph, which is a 2D projection of a 3D problem. As a result, RMP may be prone to systematic and random errors due to variation in patient positioning, and estimation of the RMP from planar images.[10–12]

3D imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI), have the potential to provide more reliable and complete assessment of hip development than planar radiographs. However, they are not without limitations. CT scans involve significant amounts of ionizing radiation, which is not acceptable in routine monitoring, MRI is expensive, and both modalities are susceptible to movement artefacts. To achieve optimal image quality using MRI or CT, a significant proportion of children would be likely to require sedation or anaesthesia. In contrast, 3D ultrasound is relatively quick and, given the non-ionizing nature, the images can be acquired safely multiple times if the patient moves during acquisition. Therefore, 3D ultrasound may provide a suitable alternative to planar radiographs for the routine surveillance of hip dysplasia.

The primary objective of this study was to perform preliminary validation of 3D ultrasound for the assessment of lateral displacement of the femoral head in children with CP. For this, a new ultrasound-based index, lateral head coverage (LHC), is defined and compared with RMP. The

DOI: 10.1111/dmcn.14647

secondary objective was to establish the intra- and inter-assessor reliability of our new index, LHC.

METHOD

Participants

Twenty-two participants (15 males, seven females), aged between 4 and 15 years, were recruited to the study. Participants were identified from paediatric orthopaedic clinics at a tertiary-level teaching hospital. The inclusion criteria stipulated that the participants must have a diagnosis of CP, be aged between 2 and 16 years, have had a 2D radiograph of the hips as part of their routine clinical management within the last 2 months, and not have undergone bony surgery to the acetabulum.

3D ultrasound assessment

Ultrasound images were acquired using either the Philips EPIQ 7 (Koninklijke Philips, Amsterdam, the Netherlands) with a 3D array probe, or the GE Voluson (GE Healthcare, Chicago, IL, USA) with a mechanical sweep probe. The depth of the scan was set between 6cm and 8cm, depending on the child's size, with a sweep angle of 60°. Each child was positioned so that they were side-lying with hips extended as close to neutral as possible. The probe was orientated parallel to the superior–inferior axis of the pelvis over the lateral aspect of the hip. To optimize image acquisition, the greater trochanter was identified, and the probe was moved posterior–superiorly to obtain an optimal view of the femoral head and lateral acetabular border (Fig. 1). Images were saved and exported in DICOM or GE.vol (GE Healthcare) format.

Slicer v4.10.1, an open source image processing software (https://www.slicer.org/),[13] was used for image analysis. The slice in the coronal plane with the maximum femoral head cross-section that also corresponded with the maximum femoral cross-section in the sagittal plane was selected for analysis (Fig. 1). A 'best fit' sphere was fitted to the femoral head and the diameter measured as an estimate of femoral head diameter. The lateral aspect of the acetabulum was also identified in the same medio–lateral image slice, and the lateral distance between the acetabulum border and the lateral aspect of the femoral head measured. The ratio of the two measurements was taken and deducted from 1 to give an estimate of the proportion of the femoral head that was covered by acetabulum. We will refer to this new index as LHC. LHC is an index for quantifying the femoral head coverage in the coronal plane, where 'e' is the distance in the lateral plane from the lateral aspect of the acetabulum to the lateral aspect of the femoral head, and 'd' is the diameter of the 'best fit' sphere.

$$LHC = 1 - \left(\frac{e}{d}\right) \times 100$$

Because of the construction of an ultrasound image it is not possible to identify the same bony landmarks as an X-ray, thus it is not possible to create a mathematically equivalent index. For this reason, LHC measures coverage of the femoral head by the acetabulum, and not the proportion of the femoral head that is uncovered, as is done in RMP.

What this paper adds

- Reliability of measuring 3D ultrasound assessment of lateral head coverage (LHC) was comparable with reported Reimer's migration percentage (RMP).
- Strong correlation was found with 3D ultrasound assessment of LHC and the clinical standard (RMP).

It was necessary to exclude some hips from the validation study if the necessary bony landmarks could not be identified in either the X-ray or the 3D ultrasound. Table 1 shows each participant and hip with details of inclusion/exclusion in this study. LHC was measured on the remaining 24 hips, which met the image inclusion criteria, by assessor 1.

Radiographs were acquired as part of the routine clinical care of each child under the standard positioning protocols used for hip surveillance. Participants were supine ensuring that the hips were as close to neutral in rotation, abduction/adduction, and flexion/extension as the children's hip ranges allowed. The radiographers were unaware of the children's participation in a research study. RMP was measured by assessor 1 and the classic method was used (i.e. using the lateral aspect of the acetabulum) (Fig. 1) because of its superior reliability.[14] All measurements were made using the image analysis package PACS Sectra IDS 7 (v21.1.5 2096; Sectra AB, Linköping, Sweden). All RMP measurements were made with an interval of at least 1 week from the corresponding ultrasound measurement.

Repeatability

Eleven of the 24 hips included in the study were selected at random to investigate the repeatability of the 3D ultrasound image analysis. Three assessors with varying experience in analysing 2D B-mode ultrasound (2mo–7y) analysed the images. Assessor 1 had 6 months experience of analysing 3D ultrasound images of the hip, while the other two assessors had no prior experience in analysing these images. The two inexperienced assessors underwent an initial training session led by assessor 1 and had an opportunity to practice, compare, and receive feedback on a training set of images before beginning the study. All study images were different to the training images. Each assessor used Slicer v4.10.1 to analyse the images. All identifying information was removed. In the first session, each image was analysed twice with the image order randomized. In the second session, at least 1 week later, each image was analysed once, resulting in each image being analysed three times by the three assessors across the two sessions. All scores were sent to the study coordinator for compilation. One image was removed from the study as the acetabular border was consistently not visible in the slice chosen for analysis by the assessors.

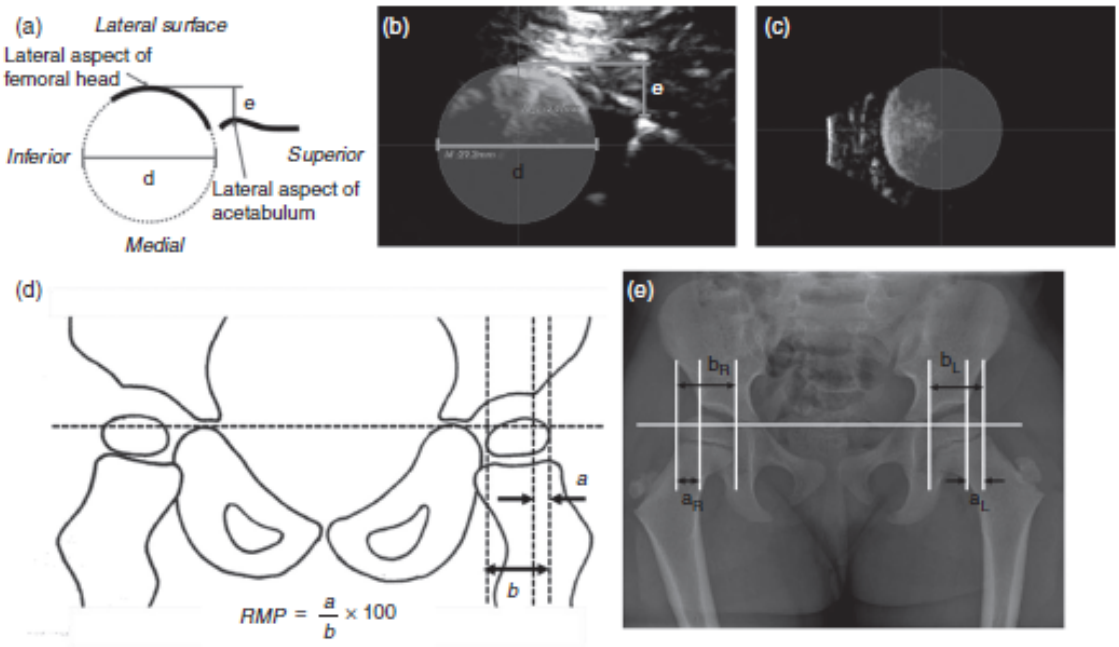


Figure 1: (a) A schematic of the coronal plane image showing the lateral aspect of the femoral head and the superior lateral aspect of the acetabulum. (b) The coronal plane slice acquired in side-lying position showing the 'best fit' sphere; 'e', distance between lateral aspect of acetabulum and lateral aspect of femoral head; and 'd', estimated femoral head diameter. (c) The sagittal plane slice again showing the 'best fit' sphere. (d) A schematic of the measurement of Reimer's migration percentage (RMP). (e) X-ray showing measurement of RMP.

Data analysis

To investigate whether the two indices were significantly different, a paired Student's *t*-test was used. The RMP and LHC should be inversely proportional to each other, as one describes the proportion of the femoral head that is not covered by the acetabulum (RMP) and the other describes the proportion of the femoral head that is covered by the acetabulum (LHC). In order to conduct the paired Student's *t*-test, 1–LHC was calculated. Pearson's rank correlation coefficient was used to assess the strength of correlation between the RMP and LHC. SPSS v26 (IBM Corp., Armonk, NY, USA) was used for the statistical analyses.

The inter- and intra-assessor repeatabilities of the LHC measurements were investigated using intraclass correlation coefficients (3,1) (two-way mixed, single measures). SPSS v26 was used to compute the intra-assessor intraclass correlation coefficient (3,1) (two-way mixed, single measures) using 90 measurements (3 assessors×10 images×3 repeats). The inter-assessor intraclass correlation coefficient was calculated using the first of the repeat images from each assessor (a total of 30 measurements, 10 per assessor).

To investigate potential bias between sessions, the mean of the two measurements from the first session for each assessor was deducted from the second session measurements. The means and standard errors of the mean (SEM) differences were calculated to allow calculation of the upper and lower bias limits.

To aid the establishment of the potential clinical utility of LHC, the minimal detectable difference (MDD) was calculated. The MDD is the smallest change in two measurements that can confidently (95% confidence intervals [CIs]) be taken as a true difference.

Ethical consent

Written consent was sought from the participants' parents or legal guardians. The study was approved by the Health Research Authority and the national research ethics committee (Wales REC 7 study number 17/WA/0093).

RESULTS

Exclusions

Table 1 lists all recruits and details the rationale for exclusion from the validation study. Out of the 44 hips available, 24 were included in the study (age range 5y 7mo–13y 4mo, mean age 9y 5mo, standard deviation 2y 6mo). The most frequent reason for exclusion was that the ultrasound probe was not correctly orientated at the point of image acquisition. An amendment to the image acquisition protocol was made after participant 9, to ensure that the same view was reliably acquired and recorded. After this amendment, only two hips were excluded because of an ultrasound acquisition issue. Two further comparisons were excluded due to poor X-ray contrast rendering the border of the acetabulum

Table 1: Inclusions/exclusions for each participant and each hip within the study

Participant	Sex	Left hip	Right hip
1	Male	No orientation data recorded	No orientation data recorded
2	Male	No orientation data recorded	No orientation data recorded
3	Male	Included	Included
4	Female	No imaging with probe in correct orientation acquired	No imaging with probe in correct orientation acquired
5	Male	No acetabulum visualized	Included
6	Male	Included	No acetabulum visualized
7	Male	No imaging with probe in correct orientation acquired	No imaging with probe in correct orientation acquired
8	Female	No imaging with probe in correct orientation acquired	No imaging with probe in correct orientation acquired
9	Male	No imaging with probe in correct orientation acquired	No imaging with probe in correct orientation acquired
10	Female	Included	Included
11	Male	Included	Included
12	Male	Included	Included
13	Male	X-ray not interpretable	X-ray not interpretable
14	Female	Included	Included
15	Female	Included	Included
16	Female	Child did not tolerate	Child did not tolerate
17	Female	Included	Included
18	Male	Included	Acetabulum not visualized in slice chosen as maximal cross-section of femoral head
19	Male	Included	Included
20	Male	Included	Included
21	Male	Included	Included
22	Male	Included	Acetabulum not visualized in slice chosen as maximal cross-section of femoral head

undefinable. In two instances, the child did not tolerate the ultrasound examination.

Validation study

RMP was not significantly different to 1–LHC (Student's *t*-test [df 23]=–0.494; p=0.626). Figure 2 shows the relationship between the RMP and LHC. There was a strong correlation between RMP and LHC, with a Pearson's rank correlation coefficient of –0.86 (p<0.001).

Reliability

No bias between sessions was detected for any of the assessors. The SDs of the averaged measurements between the assessors ranged from 0.47% for image number 3, with best agreement, to 6.99% (Fig. 3). Interclass reliability (intradass correlation coefficient [3,1] (two-way mixed, single measures)) was 0.973 (95% CI 0.925–0.998), and a corresponding SEM of 3.6% and MDD 10%. Intraclass reliability was 0.982 (95% CI 0.967–0.991).

DISCUSSION

This paper describes the development of a new index, LHC, for the quantification of the lateral coverage of the femoral head by the acetabulum. LHC demonstrated a strong correlation with RMP, with 74% of the variation in LHC being explained by RMP. Further, both the inter-

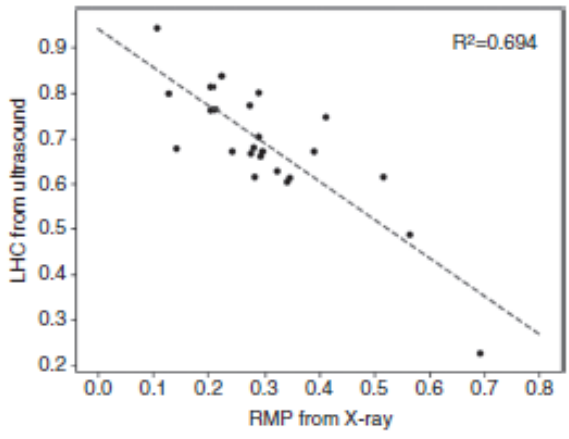


Figure 2: F A comparison of Reimer's migration percentage (RMP), measured by X-ray, with lateral head coverage (LHC) measurement measured by assessor 1 for 24 hips.

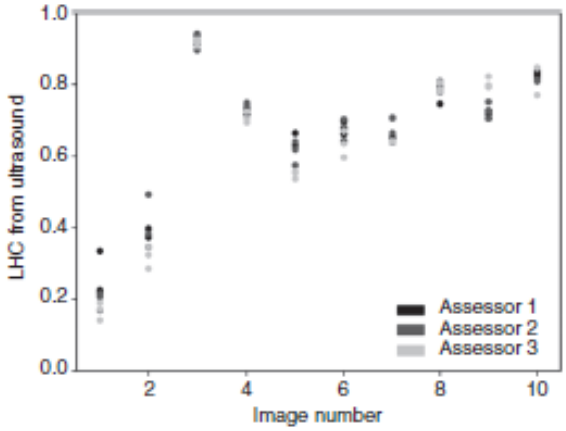


Figure 3: F Lateral head coverage (LHC) measured three times by three assessors for each of the 10 image volumes. Assessor 3 was the most reliable with a standard error of the mean (SEM) of 2.39%. Assessors 1 and 2 were very similar with SEMs of 2.91% and 2.97% respectively.

and intra-assessor reliability of LHC were excellent, and similar to that reported for RMP.[10,12,15] The SEMs and corresponding MDDs for LHC were also comparable with those reported for RMP (SEM range 2.98–3.9%, MDD=8.3–11.5%).[10,12]

LHC is a measure that is simply constructed using both the sagittal and coronal planes of the ultrasound volume. Ensuring that the maximal cross-sectional area is found in two orthogonal slices removes a source of error to which 2D radiographs are prone. LHC gives an indication of the percentage of lateral coverage of the femoral head by the acetabulum, in contrast to RMP, which measures the percentage of femoral head that is uncovered. We considered using an index analogous to RMP but, to minimize confusion between our prospective index and RMP, we chose to create an index describing the coverage of the femoral head because LHC and RMP are informed by different bony landmarks, and are computed by different methods.

This is the first study to use 3D ultrasound to assess hip development in children with CP. There have been only a small number of studies that have used 2D ultrasound imaging in this population. Smigovec et al.[16] visualized hips in children with severe CP (GMFCS level IV or V) using 2D B-mode ultrasound. The scanning technique described here is an adaptation of the method used by Smigovec et al.[16] Smigovec and colleagues[16] reported encouraging results, discriminating between measurements above and below a threshold RMP with greater than 90% sensitivity and specificity. Before their work, Tegnander and Terjesen[17,18] investigated the feasibility and reliability of using 2D ultrasound to assess fully ossified hips in children above 2 years of age. Initially, they looked at 'normal hips' (i.e. children with no previous hip pathology) and concluded that the required bony landmarks could be visualized to measure the coverage of the femoral head by the acetabulum. They proposed normal limits for coverage (by their index) depending on age.

Previous efforts to investigate the use of ultrasound in visualizing hip development in older children and specifically children with CP may well have been stalled by limitations related to interoperator variance.[19] In contrast, 3D ultrasound is proving to be an accurate and reliable tool in the morphological evaluation of the musculoskeletal system,[20] specifically in soft tissue imaging,[20–22] but few studies to date have analysed proximal femoral or hip geometry. Passmore et al.[23] used freehand 3D ultrasound to measure femoral neck anteversion angle, comparing results to those obtained from MRI. The correlation was high (Pearson's rank correlation coefficient=0.94) with an average difference of 1.8° between the imaging modalities across the 10 individuals. 3D ultrasound was found to have repeatability coefficient of 3.7°, which was comparable with that of MRI, which was reported as 3.1°.

Limitations

There were a significant number of exclusions from this study, largely due to inexperience in the ultrasound image acquisition of the hip at the start of this study. Such early images, acquired with the probe not orientated in line with the superior–inferior axis of the pelvis, were excluded.

Hip dysplasia is a 3D maldevelopment of the hip and, as such, would ideally be monitored by 3D imaging. Traditional 3D imaging modalities (CT and MRI) are not valid options for routine monitoring of hip development in children with CP because of increased exposure to ionizing radiation (CT), the expense of the investigations, and the requirement for sedation or anaesthesia in younger children. When making clinical decisions, orthopaedic surgeons and others need to be aware of the limitations of measurements from planar radiographs in the assessment of the hip.

LHC was compared with RMP, as RMP is the clinical standard for routinely monitoring hip development in children with CP. This was a pragmatic decision for recruitment purposes. However, RMP is not a criterion standard measurement, and it is not clear what contribution errors in RMP measurement have on the comparison between LHC and RMP.

For more complete validation, LHC should be compared with a similar index derived from 3D imaging, where

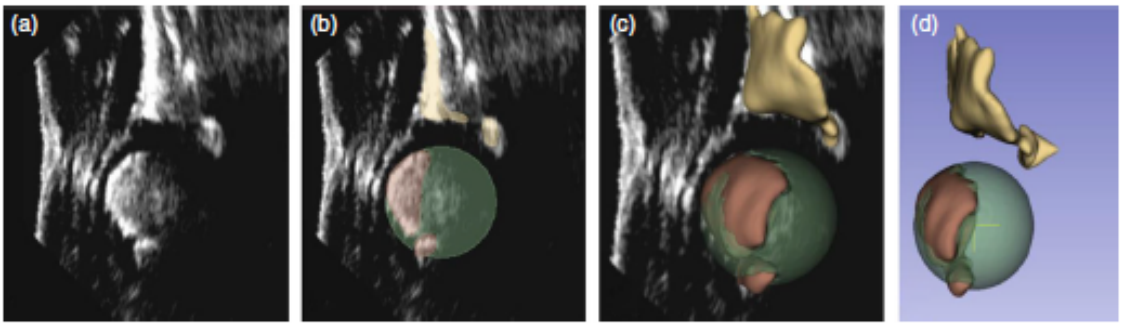


Figure 4: All images are of the same left hip from a 4-y-old male. In all images, the top of the image is anatomically superior and lateral on the left side, and medial on the right side. (a) A 2D coronal plane slice at maximal cross-sectional area of the femoral head. (b) 2D coronal plane slice with surface rendering superimposed. (c) A 2D coronal slice with 3D surface rendering. (d) A coronal view of the 3D surface render, showing the 'best fit' sphere superimposed over the femoral head.

error sources could be minimized and the comparative measurements could be considered a true criterion standard. Gose et al.[24] compared a CT-derived index to RMP and reported a strong correlation (Pearson's rank correlation coefficient=0.85, $p<0.001$) between the measurements, comparable with the agreement found in our study. Figure 4 shows a 3D render of the left hip from a child with CP. Using 3D ultrasound, detailed hip morphology can be visualized. However, to validate the observations, it would be of interest to compare 3D ultrasound to a criterion standard 3D imaging modality such as MRI or CT.

RMP and LHC rely on the identification of the lateral border of the acetabulum and a measurement of femoral head diameter/width. However, the two modalities construct images in different ways. Planar radiographs are projection images showing areas of high and low absorption of the X-rays as they pass through the object from source to receiver, which allows for high contrast between bone (highly absorbent) and surrounding soft tissues (less absorbent), normally resulting in clear 2D imaging of hip morphology. The width of the femoral head is measured at the level of Hilgenreiner's line. Ultrasound images are constructed from the reflections of the soundwaves at borders between different tissues. Bony surfaces are highly reflective to these soundwaves and, as such, ultrasound cannot visualize structures that sit close to a bony surface. In order to get an estimate of femoral head size, a sphere of 'best fit' was fitted to the lateral curvature of the femoral head in both the sagittal and coronal planes (Fig. 1), and the diameter of the sphere was taken as an estimate of femoral head size. When comparing the estimates of femoral head diameter from ultrasound to those derived by X-rays we did not detect a systematic difference between the measurements. The absolute percentage error in the measurements was computed to be 10.1%. This error may be due to the errors associated with measurements made on both X-ray and ultrasound images.

Clinical implications

The use of ultrasound to evaluate hips in young infants has transformed the screening of developmental hip dysplasia.[25] Developmental hip dysplasia is an umbrella term used to describe abnormal positioning of the femoral head and acetabulum in otherwise typically developing infants. Ultrasound lends itself well to imaging of the hip in the very young, as the hip has not ossified and therefore sound waves are able to partially penetrate through the hip joint, allowing visualization of the acetabulum. As the hip ossifies, it becomes impossible to get clear images of the joint. However, as this study has shown, it is still possible to visualize significant anatomical landmarks and make measurements of hip geometry that may have diagnostic value.

Using 3D ultrasound imaging would allow for more frequent and repeated assessments to be performed, because ultrasound is a non-ionizing imaging modality. Ultrasound imaging would also allow for the hip to be imaged in different positions, providing further information about hip development that cannot currently be collected from single radiographs. Repeated measurements in the same position would permit greater confidence in the estimation of hip displacement. Screening programmes often do not have frequent monitoring for children with CP who are less severely affected as they are less at risk of hip displacement. Depending on the programme, individuals may be discharged after a single 'normal' radiograph or receive a further X-ray at around the age of 8 years (after which very few hips go on to dislocate).[6,8,26] Kentish et al.[27] reviewed 1115 children who had been engaged in their hip screening programme. Of these, 28% had RMP of greater than 30%. In this group with high RMPs, 16% were in GMFCS levels I or II. Using a non-ionizing imaging modality such as 3D ultrasound would permit safe continued monitoring in the more able group.

CONCLUSION

This paper presents an initial validation of the use of 3D ultrasound in monitoring hip development in children with CP. The results show that LHC is comparable with RMP in estimating hip dysplasia, with similar levels of reliability to those reported for X-ray. With the potential to increase assessment frequency, the 3D ultrasound assessment technique could, as a minimum, provide a non-ionizing alternative for the monitoring of hip dysplasia in children with CP. It is also likely that the additional structures and views that can be imaged with ultrasound compared with a 2D radiograph could provide valuable information on hip management for individuals with CP. Further investigations are required to appreciate the full potential of 3D ultrasound in the monitoring of hip dysplasia in children with CP.

ACKNOWLEDGEMENTS

Thanks to Dr Anne McNee, Dr Robert Eckersley, and Professor Lionel Tarassenko for their guidance and advice. This work was funded by Action Medical Research (UK), a registered charity in England and Wales (208701) and in Scotland (SC039284), and supported by the Wellcome/Engineering and Physical Sciences Research Council Centre for Medical Engineering (WT 203148/Z/16/Z).

REFERENCES

1. Scrutton D, Baird G. Surveillance measures of the hips of children with bilateral cerebral palsy. *Arch Dis Child* 1997; 76: 381–4.
2. Soo B, Howard JJ, Boyd RN, et al. Hip displacement in cerebral palsy. *J Bone Joint Surg Am* 2006; 88: 121–9.
3. Cooke PH, Cole WG, Carey RPL. Dislocation of the hip in cerebral palsy. Natural history and predictability. *J Bone Joint Surg Br* 1989; 71: 441–6.

4. Hägglund G, Alriksson-Schmidt A, Lauge-Pedersen H, Rodby-Bousquet E, Wagner P, Westbom L. Prevention of dislocation of the hip in children with cerebral palsy: 20-year results of a population-based prevention programme. *Bone Joint J* 2014; 96: 1546–52.
5. Larnert P, Risto O, Hägglund G, Wagner P. Hip displacement in relation to age and gross motor function in children with cerebral palsy. *J Child Orthop* 2014; 8: 129–34.
6. Miller F, Bagg MR. Age and migration percentage as risk factors for progression in spastic hip disease. *Dev Med Child Neurol* 1995; 37: 449–55.
7. Hermansson M, Hägglund G, Riad J, Rodby-Bousquet E, Wagner P. Prediction of hip displacement in children with cerebral palsy: development of the CPUP hip score. *Bone Joint J* 2015; 97: 1441–4.
8. Hägglund G, Lauge-Pedersen H, Wagner P. Characteristics of children with hip displacement in cerebral palsy. *BMC Musculoskelet Disord* 2007; 8: 101.
9. Reimers J. The stability of the hip in children. A radiological study of the results of muscle surgery in cerebral palsy. *Acta Orthop Scand Suppl* 1980; 184: 1–100.
10. Parrott J, Boyd RN, Dobson F, et al. Hip displacement in spastic cerebral palsy: repeatability of radiologic measurement. *J Pediatr Orthop* 2002; 22: 660–7.
11. Cliffe L, Sharkey D, Charlesworth G, Minford J, Elliott S, Morton RE. Correct positioning for hip radiographs allows reliable measurement of hip displacement in cerebral palsy. *Dev Med Child Neurol* 2011; 53: 549–52.
12. Craven A, Pym A, Boyd RN. Reliability of radiologic measures of hip displacement in a cohort of preschool-aged children with cerebral palsy. *J Pediatr Orthop* 2014; 34: 597–602.
13. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012; 30: 1323–41.
14. Kim SM, Sim EG, Lim SG, Park ES. Reliability of hip migration index in children with cerebral palsy: the classic and modified methods. *Ann Rehabil Med* 2012; 36: 33–8.
15. Kinch K, Campbell DM, Maclean JGB, et al. How critical is patient positioning in radiographic assessment of the hip in cerebral palsy when measuring migration percentage? *J Pediatr Orthop* 2015; 35: 756–61.
16. Smigovec I, Dapic T, Trkulja V. Ultrasound screening for decentered hips in children with severe cerebral palsy: a preliminary evaluation. *Pediatr Radiol* 2014; 44: 1101–9.
17. Tegnander A, Terjesen T. Ultrasound measurements in hips of children above 2 years of age: normal variations in 232 hips. *Acta Orthop Scand* 1995; 66: 229–33.
18. Tegnander A, Terjesen T. Reliability of ultrasonography in the follow up of hip dysplasia in children above 2 years of age. *Acta Radiol* 1999; 40: 619–24.
19. Thoirs K, English C. Ultrasound measures of muscle thickness: intra-examiner reliability and influence of body position. *Clin Physiol Funct Imaging* 2009; 29: 440–6.
20. Barber L, Barrett R, Lichtwark G. Validation of a free-hand 3D ultrasound system for morphological measures of the medial gastrocnemius muscle. *J Biomech* 2009; 42: 1313–9.
21. Fry NR, Childs CR, Eve LC, Gough M, Robinson RO, Shortland AP. Accurate measurement of muscle belly length in the motion analysis laboratory: potential for the assessment of contracture. *Gait Posture* 2003; 17: 119–24.
22. Fry NR, Gough M, Shortland AP. Three-dimensional realisation of muscle morphology and architecture using ultrasound. *Gait Posture* 2004; 20: 177–82.
23. Passmore E, Pandy MG, Graham HK, Sangeux M. Measuring femoral torsion in vivo using freehand 3-D ultrasound imaging. *Ultrasound Med Biol* 2016; 42: 619–23.
24. Gose S, Sakai T, Shibata T, Murase T, Yoshikawa H, Sugamoto K. Morphometric analysis of acetabular dysplasia in cerebral palsy: three-dimensional CT study. *J Pediatr Orthop* 2009; 29: 896–902.
25. Kotlarsky P, Haber R, Bialik V, Eidelman M. Developmental dysplasia of the hip: what has changed in the last 20 years? *World J Orthop* 2015; 6: 886–901.
26. Terjesen T. The natural history of hip development in cerebral palsy. *Dev Med Child Neurol* 2012; 54: 951–7.
27. Kentish M, Wynter M, Snape N, Boyd R. Five-year outcome of state-wide hip surveillance of children and adolescents with cerebral palsy. *J Pediatr Rehabil Med* 2011; 4: 205–17.