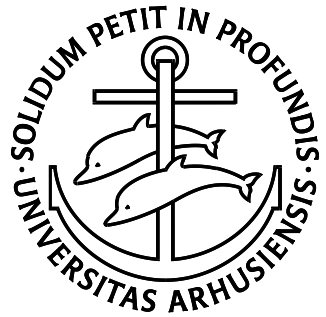
Outcomes after the calcaneal lengthening osteotomy with artificial structural bone graft in paediatric flatfoot surgery

PhD thesis

Polina Martinkevich



Faculty of Health Sciences

University of Aarhus

2016

Contact address

Polina Martinkevich, MD
Department of Children's Orthopaedics
Aarhus University Hospital
Nørrebrogade 44, Building 1C, 2nd floor
8000 Aarhus C, Denmark

+ 45 23 74 25 14
pmar@clin.au.dk
polina.martinkevich@gmail.com
www.dpor.dk

AARHUS UNIVERSITET

dpor Danish Paediatric Orthopaedic Research since 2011

Outcomes after the calcaneal lengthening osteotomy with artificial structural bone graft in paediatric flatfoot surgery

PhD thesis

Polina Martinkevich

Faculty of Health Sciences

University of Aarhus

Supervisors

Bjarne Møller-Madsen, Professor, MD, DMSc (Main supervisor)
Department of Children's Orthopaedics
Aarhus University Hospital

Ole Rahbek, Associate Professor, MD, PhD
Department of Children's Orthopaedics
Aarhus University Hospital

Maiken Stilling, Associate Professor, MD, PhD
Department of Clinical Medicine, Aarhus University
Department of Orthopaedics, University Clinic of Hand, Hip and Knee Surgery,
Regional Hospital Holstebro, Hospital Unit West

Evaluation committee

Unni Narayanan, Ass. Professor, MD, PhD, FRCS
Divisions of Orthopaedic Sugery & Child Health Evaluative Sciences Director, Pediatric Orthopaedic Fellowship, The Hospital for Sick Children, University of Toronto, Canada

Andrew Morris Wainwrigth
BSc, MB, ChB, FRCS, MEd (SE)
Oxford University Hospital, UK

Chairman and moderator
Torben Bæk Hansen, Prof, MD, PhD
Holstebro Hospital, Denmark

Acknowledgements

This dissertation is composed of three studies conducted at Aarhus University Hospital in the timespan 2013-2016. It has been an incredible journey beyond my imagination, which would not have been the same without the people who went along.

First of all, I thank my supervisors. Each of you is a pillar of the project in your very unique ways. Bjarne, your infinite optimism encouraged me to keep my head up high. Thank you for believing in me and for reinforcing the mantra that nothing is impossible. I am grateful for the faith you have shown in me. Ole, thank you for always being there to keep any process through completion. Thank you for always being present in any given situation. Maiken, your invaluable research experience and attention to detail made my work dogma this: Accuracy and precision of any method is a virtue. Thank you, for setting the expectation bar high and always encouraging me to work hard for something to make it happen.

I thank my research fellows at the tiny *Penthouse* for always being there to manage all the small practical real-life issues, that comes in the way.

The project would not have been possible to conduct without the invaluable effort from the RSA team, Rikke, Lone, Mark, Lars, Lis, Peter Bo; the Children's Orthopaedic Department team: Ann-Louise, Lise, Annie, Jeannie, Lene Sloth and colleagues and Gitte Nyhave. Thank you Michel, for spending your busy afternoons enlightening me on radiology

Thanks to my dear friends. Marie, for regularly checking up with me, and being the solid rock I needed during the difficult days of the writing phase.

Taun, you inspire me more than you realise. By taking chances, by striving to always be present in everything you do, and always seeking new perspectives makes you grow. You reinforced that in me.

Thomas Johansen. I owe you for introducing *"The Rule of π"* into my life - everything you plan always schedule times 3.14.

My research time spent abroad made me encounter inspirering personalities. Professor Unni Narayanan. Your central question *"Why do we do it?"* is always imprinted in my mind. Professor Deborah Eastwood, You taught me the value of taking time listening to the patient because if the patient's life will not truly benefit from surgery, then surgery has no place. Christopher Morris. Thanks for seeing a light in me and helping me to understand the value of PROMs and patient-healthcare provider relations.

My dear Yogi tribe. The journey we went on is the most incredible, speechless adventure I have ever been on. It radically changed my life. And each and everyone of you were a part of it.

Especially, I thank you Morten, for reminding me to be present in everything I do, for your endless support and your unconditional love.

The work of this thesis was supported by The Central Denmark Region Research Fund, The Graduate School of Health, Ceramisys Ltd, Lægeforeningens Forsikringsfond 1981, The Orthopedic Research Fund, The DOS Fund, Eiler and Tatiana Filtenborgs Fellowship Fund, Oticon Foundation and Knud Højgaard Foundation.

Polina Martinkevich
Aarhus, July 2016

This thesis is based on the following papers

Paper I

Validation of the translated Oxford ankle foot questionnaire in 82 Danish children aged between five to 16 years

P. Martinkevich, B. Møller-Madsen, M. Gottliebsen, L. Kjeldgaard Pedersen, O. Rahbek. Bone Joint J: 2015;95-b:420-6

Paper II

Precise and feasible measurements of lateral calcaneal lengthening osteotomies with radiostereometric analysis (RSA) in cadaver feet

P. Martinkevich, O. Rahbek, B. Møller-Madsen, K. Søballe, M. Stilling. Bone Joint Res 2015;4:78-83

Paper III

Structural HATCP bone graft versus tricortical iliac crest autograft in paediatric calcaneal lengthening osteotomies. Interim results from a randomised, controlled, noninferiority study

P. Martinkevich, O. Rahbek, M. Stilling, L.K. Pedersen, M. Gottliebsen, K.Søballe, B. Møller-Madsen. Bone and Joint Journal. Accepted for publication the 15th July 2016.

Abbreviations

AI	Arch Index
AUH	Aarhus University Hospital
AUTO	Autologous
CCJ	Calcaneal-cuboid joint
CI	Confidence interval
CLO	Calcaneal lengthening osteotomy
CoNS	Coagulase-negative staphylococci
dCF	Distal calcaneal fragment
HATCP	Hydroxyapatite-tricalciumphosphate
HRQoL	Healt-related quality of life
ICBG	Iliac crest bone graft
LOA	Limits of agreement
L	Liters
lp	Line pairs
mm	Milimeters
OxAFQ	Oxford Ankle Foot Questionnaire
pCF	Proximal calcaneal fragment
PPV	Pes planovalgus
PROM	Patient reported outcome measure
RSA	Radiostereometric analysis; radiostereometry
SA	Staphylococcus Areus
SD	Standard deviation

Contents

English summary

The calcaneal lengthening osteomy (CLO) is among the most widely used surgical techniques for correction of pes planovalgus deformities in paediatric patients. The choice of structural bone graft material has been debated ever since the introduction of the operation. Most studies have focused on the feasibility of allograft as an alternative to tricortical iliac crest autograft. Over the past decades, there has been a growing interest in bone graft substitutes, especially calcium ceramics owing to their resemblance to native bone. If calcium ceramic could be proven viable as a structural bone graft, it would be a promising alternative to iliac crest autograft for CLO. When assessing the efficacy of a structural bone graft, two major aspects need to be determined: The mechanical properties of the bone graft and the clinical outcome.

Radiostereometric analysis (RSA) is a valuable tool in the measurement of migration of implants.[1] It has not yet been used to assess bone incorporation and osteotomy stability in paediatric foot surgery.

Recent trends in using Patient Reported Outcome Measures (PROMs) to measure patient health have resulted in development of tailored PROMs, seeking to ascertain patients' views on how their symptoms affect their daily lives. The specific English Oxford Foot Ankle Questionnaire (OxAFQ) has been proven valid in paediatric patients and their parents. For Danish paediatric patients with foot-ankle deformities including pes planovalgus, such a PROM would be of great clinical value to aid the assessment of how children are affected by their foot deformity.

This dissertation seeks to investigate the feasibility and efficacy of a novel structural calcium ceramic for paediatric CLO. The osteotomy stability is assessed with RSA and is thus a surrogate measure of the structural bone graft durability. Secondary, the outcome will be assessed by the OxAFQ to assess the patient's perspective on the outcome.

The first aim *(Study I)* of this thesis was to translate the English version of OxAFQ into Danish with subsequent validation. The second aim *(Study II)* was to assess the precision and feasibility of RSA measurements for CLO. The third and central aim *(Study III)* was to investigate the noninferiority of structural hydroxyapatite-tricalciumphosphate (HATCP) to tricortical iliac crest autograft in CLO for peadiatric pes planovalgus by using RSA as the primary measurement instrument.

Study I. The English version of OxAFQ (child and teenager, and proxy version) was translated into Danish. The subsequent validation included reliability measurements (test-retest, child-parent, internal consistency), construct validity and feasibility (number of

missing items, time to completion and floor and ceiling effects). The generic child health questionnaire was used for comparison. Overall the OxAFQ was valid and feasible in assessment of health-related quality of life in children with a variety of foot/ankle disorders.

Study II. Precision and feasibility of RSA measurements for CLO were tested in three fixed adult cadavers (six feet) with the primary focus on the stability of the osteotomy and secondary the calcaneal-cuboid joint (CCJ). Precision was good for translations at the osteotomy site and at the CCJ. As expected, rotations were less precise, but without clinical significance. RSA proved to be feasible for CLO in fixed cadaver feet, and is expected to be more feasible in paediatric feet.

Study III. The study was designed as a randomised controlled noninferiority trial. The primary outcome measure was x-translation of the distal (anterior) calcaneal fragment with reference to the proximal (posterior) calcaneal fragment measured by RSA. If there was truly no difference between the new and the standard treatment, 20 patients would be required to be 90% sure that the upper limit of a one-sided 95% CI would be above a noninferiority limit of 2 mm. For safety reasons an interim-analysis was embedded into to the trial. Data from ten patients with six months RSA follow-up were analysed with respect to the primary outcome measure considering both sides of a 99.8% CI. The interim results of *Study III* are reported in this dissertation. The report includes eleven patients (11 feet). One patient was excluded from further follow-up after eight weeks due to infection of the graft material (HATCP), which necessitated reoperation. With six patient in the HATCP group and five patients in the AUTO group, a difference of 1.97 mm (99.8%CI: -1.65; 5.60) was found in the x-translation of the distal calcaneal fragment with reference to the proximal calcaneal fragment (HATCP graft minus AUTO graft). In general, the AUTO group obtained stability of the osteotomy at six weeks follow-up, while the HATCP group stabilised at six months.

This is the first clinical study to compare the structural durability of iliac crest autograft with HATCP for CLO in paediatric patients by using RSA. The results suggest that HATCP is of limited value as a structural bone graft in unfixed CLOs compared to tricortical iliac crest autograft.

Danish summary

Forlængelse af calcaneus er en populær kirurgisk teknik til at korrektion af fleksible platfod hos børn. Siden denne operation blev introduceret, har der været et stort fokus på at finde den optimale strukturelle knoglegraft. Forskning inden for børneortopædi har sammenlignet indhelingen af autolog trikortikal crista iliaca graft med allogen knogle, men resultaterne har været inkonsistente. I løbet af de seneste årtier har der været en stigende interesse for artificielle knoglegraft substitutter herunder calciumkeramik grundet materialets ligheder med nativ knogle. Der foreligger endnu ingen randomiserede studier, der sammenligner de strukturelle egenskaber af calciumkeramik og trikortikal autograft i calcaneus-forlængelsesosteotomier hos børn. I vurderingen af strukturel knoglegraft bør to vigtige aspekter bestemmes: Stabiliteten af strukturel knoglegraft og det kliniske udfald.

Stereorøntgen er en lovende metode til at detektere migration af implantater, især inden for knæ-og hoftealloplastisk, men det har endnu ikke været anvendt på calcaneusforlængelsesostetomier hos børn.

De seneste tendenser inden for anvendelsen af patientrapporterede outcome mål, såkaldte PROMs, har medført en udviklingen i at skræddersy PROMs, der har til formål at beskrive patientens perspektiv på egne symptomer, samt i hvilket omfang disse påvirker almindelige dagligdagsaktiviteter. Det regions/dimensionsspecifikke "*Oxford Ankle Foot Questionnaire*", OxAFQ, er vist at være et validt instrument til at skelne mellem sværhedsgrader af fod-ankellidelser hos børn og unge samt til at detektere forandringer over tid. Vi mangler et spørgeskema af dette format til danske børn med fod-ankellidelser herunder fleksibel platfod, til at hjælpe ortopædkirurger med at vurdere og dokumentere i hvilket omfang børn er indskrænkede i deres funktionsniveau som følge af deres tilstand.

Det overordnede formål i denne afhandling er at undersøge anvendeligheden og effektiviteten af strukturel calciumkeramik hos børn med svær fleksibel platfod. Effektiviteten vurderes ud fra materialets evne til at stabilisere osteotomien målt med stereorøntgen samt det kliniske udfald målt med OxAFQ. Det første delmål i thesen (studie I) var at oversætte den engelske version af OxAFQ til dansk med efterfølgende validering. Det andet delmål (studie II) var at bestemme præcisionen og anvendeligheden af stereorøntgen i calcaneusforlængelsesostetomier. Det tredie og centrale delmål (studie III) var at undersøge om strukturel hydroxyapatittricalciumphosphat (HATCP) var noninferiort sammenlignet med trikortikal crista iliaca autograft i calcaneusforlængelser hos børn med svær fleksibel platfod målt med stereorøntgen.

Studie I. Den engelske version af OxAFQ (barn og teenager samt forældre (proxy) version) blev oversat til dansk med efterfølgende validering. Valideringen inkluderede reliabilitet (test-retest, internal consistency ("homogenitet")), anvendeligehed (respons, udfyldningstid, "loft og gulveffekt") og konstruktionsvaliditet. Det generiske Child Health Questionnaire i dansk version anvendtes til hypotesetestning i forbindelse med at bestemme konstruktionsvaliditet. Generalt fandt vi, at OxAF spørgeskemaet er let at anvende og validt til at vurdere indflydelsen af fod-ankelproblemer på børn og unges dagligdag.

Studie II. Præcision og anvendelighed af stereorøntgenmålinger blev testet i tre kadavere (seks fødder) med fokus primært på stabiliteten af calcaneusforlængelsesosteotomien og sekundært på calcaneo-cuboidalleddet. Præcisionen var god ved translationsmålinger, og mindre god ved rotationsmålinger. Dette er forventeligt i små led men uden klinisk relevans for det aktuelle projekt.
Stereorøntgen let anvendeligt til at undersøge calcaneusosteotomier i fikserede kadaverfødder. Derfor anses det også til at være egnet til at måle på calcaneusosteotomier i børnefødder.
Studie III. Studiet blev designed som et randomiseret kontrolleret noninferioritetsstudie. Det primære effektmål var x-translationen af det distale (forreste) calcaneusfragme i forhold til det proximale (bageste) fragme målt med stereorøntgen i ufikserede calcaneusforlængelsesosteotomier hos børn med smertefuld fleksibel platfod som udtryk for stabiliteten af forlængelsesosteotomien.

For at være 90 % sikker på at den øvre grænse i et én-sidet 95% sikkerhedsinterval vil være under en non-inferioritetsmargin på 2 mm, såfremt der ikke er nogen forskel mellem de to behandlinger, kræves 10 patienter i hver gruppe. Af sikkerhedsmæssige årsager indsattes en interim analyse, når 10 patienter, 5 i hver gruppe, havde gennemgået 6 måneders RSA opfølgning. X-translationen blev målt med et to-sidet 99.8% sikkerhedsinterval. Interimrapporten udgør *Studie III* i denne afhandling. 11 patienter (11 fødder) er inkluderet i denne rapport. Èn patient blev ekskluderet fra ydereligere undersøgelser efter 8 uger, grundet infektion af HATCP materialet, der nødvendiggjorde reoperation. Med 6 patienter i HATCP gruppen og 5 patienter i AUTO gruppen, var det gennemsnitlige tab af forlængelse 1.97 mm (99.8%CI: -1.65;5.60) i HATCP-gruppen i forhold til AUTO-gruppen. Generelt opnåede AUTO-gruppen stabilitet af ostetomien 6 uger efter operationen, hvorimod HATCP gruppen opnåede stabilitet 6 måneder efter operationen.

Dette er det første kliniske randomiserede studie, der gør brug af stereorøntgen til at sammenligne stabiliteten af strukturel HATCP mod trikortikal crista iliaca autograft i calcaneusforlængelsesostetomoier hos børn med svær fleksibel platfod. Resultaterne tyder på, at HATCP har begrænset værdi som en struktuel graft i calcaneusosteotomier, hvor der ikke anvendes fiksation af graftmaterialet.

"As the geometer his mind applies
To square the circle, nor for all his wit
Finds the right formula, however he tries"

Dante Alighieri
Paradise, canto XXXII

Introduction

Flatfoot surgery

Why we do it, how we do it and how we can measure the outcome

Flatfeet are among the most frequent clinical presentations of the foot in the paediatric orthopaedics outpatient clinics.[2] Although the literature remains diverse on the terminology and definition of a flat foot,[3] there is consensus on the following components of the deformity: The medial longitudinal arch is flattened, the heel is positioned in valgus and the lateral column appears shorter than the medial column (Figure I).[2] In this dissertation the above-described clinical presentation will be referred to as pes planovalgus.

Pes planovalgus (PPV) can be classified into two categories; rigid and flexible (with or without pain). The rigid PPV, defined by restricted subtalar joint motion, is typically associated with an underlying pathologic condition (such as a neuromuscular disease or tarsal coalition). It typically requires a different intervention than the flexible PPV, and is not within the scope of this thesis.

The majority of flatfeet are physiological and symptomless and they usually disappear by the age of 8 to 10 years.[2] The non-physiological PPV is typically associated with pain localised under the talar head during weight-bearing. If left untreated, there is a risk of impairment of normal gait development. Anatomically the calcaneus is in a valgus-position. The talar head is plantarflexed. The navicular bone is abducted (laterally translated) and dorsiflexed. This creates a convexity in the medial border of the foot with the talar head uncovered, and the medial longitudinal arch flattened.[4] The forefoot is supinated with relation to the hindfoot and the forefoot/midfoot can be either abducted or neutral.

Typical conservative treatment options for flexible PPV include physiotherapy, shoe wear modifications and orthotics. The efficiency of shoe wear and orthotics is debatable[5] and will not be discussed futher in this thesis. When conservative treatment fails and the patients are limited within their daily lives due to foot pain, surgery may be the final option. A common procedure for correction of symptomatic flexible PPV is the calcaneal lengthening osteotomy (CLO).

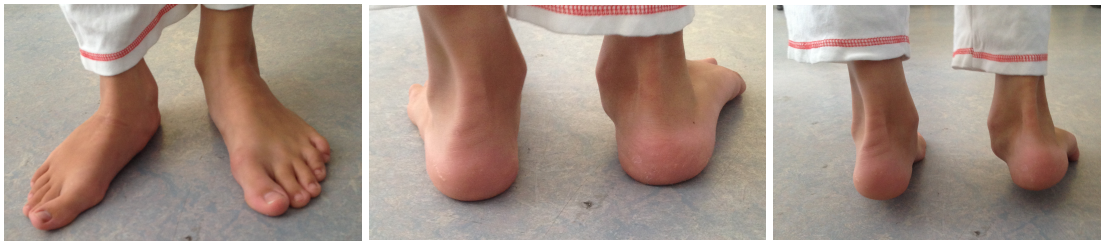


Figure I. Clinical photo of a PPV. Midfoot abduction (left), hindfoot valgus with "too many toes" sign (middle). In double-heel rise the hindfoot valgus converts into varus apperance in the flexible hindfoot (right).

Calcaneal lengthening osteotomy

The CLO is among the surgical techniques encompassed by the umbrealla term "lateral column lengthening".[6] The lateral column comprises the articulation of the anterior facet of the calcaneus with the cuboid, the 4th and the 5th tarsometatarsal joints. Any surgical procedure performed in the lateral column is typically aimed towards correction of a PPV deformity. The goal is to alleviate the patient's foot pain, thus improving the planter loading pattern and preventing early degenerative changes in the tarsal joints.

The CLO was originally performed by Evans in 1959, who later described it as a treatment for calcaneo-valgus deformities.[7] Mosca elaborated on the technique.[8] The surgical procedure corrects the paediatric PPV without interfering with the calcaneal growth plates or sacrifying the subtalar joint mobility.[9] This makes it a very popular technique for correction of the paediatric PPV. The concept of the CLO is to lengthen the calcaneus by insertion of a trapeziodal-shaped bone graft, thereby correcting all the elements of a hindfoot valgus deformity. The hindfoot valgus is reduced, the supinated forefoot is pronated and the medial longitudinal arch is restored.[10,11] Often a painful PPV is accompanied by a tightness of the Achilles tendon or the gastrocnemii muscles. This should be determined intra-operatively by the Silfverskiöld test.[11] Typically, ankle equinus is uncovered following the CLO, as the subtalar eversion is reduced.[8,12] Another important issue to consider intraoperatively, is the forefoot supination. Usually it is spontaneusly corrected following graft-insertion, but in case of a rigidity it should be adressed by a medial cuneiform osteotomy.[11]

Mosca recommended pinning the osteotomy and/or the calcaneo-cuboid joint (CCJ) if there was a concern regarding stability of the bone graft or subluxation of the distal calcaneal bone fragment.[8] Mosca used trapezoidal-shaped tricortical iliac crest bone graft, which could be either autologous or allogenous.

Over time, research has been conducted to investigate how graft type and and fixation would affect the final outcome.[6,12–19] However, the methodological weaknesses of these studies

impair the validity of the study results. One of the greatest challenges has been to determine which bone graft material would provide superior structural durability. In the following sections, I will review why.

Current bone grafting techniques in the CLO

In the history of the CLO different bone grafting techniques have been used, typically dictated by the surgeon's preference.[6] Donor site could differ, but the choice of graft type offered was to be either autologous or allogenous.[4] For many years autologous bone grafts have been considered the gold standard owing to their osteoconductivity, osteogenicity and osteoinductivity as well as mechanical stability and immunological tolerance.[20–22] The preferred structural bone graft at our institution is trapezoid-shaped autologous tricortical iliac crest bone. However, harvest of autologous bone graft is associated with donor-site morbidity. While there is no published literature review specific on donor site morbidity of tricortical iliac crest bone grafting in the paediatric population, the literature is abundent with various papers on the morbidity reported in adults.[23–37] Typically reported observations include pain (acute and chronic), sensory disturbances, hypertrophic hip scar, infection, haematoma/seroma, nerve injury. Miscellaneous rare complications have been reported.[36] In paediatric orthopaedics, it is well-known, that the pain associated with autograft harvesting often predominates the pain from the primary surgical site.[32] For many years the donor site related morbidity has led surgeons and researchers into a quest for the ideal structural bone graft. Allograft has been widely used in CLOs with reported good short-[16,38–41] and long-term results.[42] It can be obtained from bone tissue banks, it has osteoconductive potential and the patient is spared the donor-site morbidity. Albeit a suitable and popular bone graft, it has several drawbacks. Osteoinduction and mechanical stability are compromised as a result of the preparation.[43] Although the modern donor bone procurement and processing technology is designed to significantly reduce the risk of known pathogens, the concern with transmission of unknown pathogens and the risk of immunological rejection in children remains. Other limitations include high costs and limited availability of tricortical iliac crest. Commercially produced structural iliac allograft materials has been introduced to the market.[40] However, comparative scientific data on the clinical efficacy of these materials are not available. In addition ethics and price must be considered.

In paediatric foot surgery, the superior graft type has not been established by randomised trials. A further limitation to direct comparison of the previous results and their external validity, is the different criteria for bone incorporation.[16,19,40,41,44] Looking into adult foot surgery,

Dolan et al[45] conducted a randomised controlled trial comparing tricortical iliac crest autograft to allograft in lateral column lengthening for adult acuired PPV, and found no significant differences in the outcome. The primary endpoint was bone incorporation assessed by *"cortical or trabecular bridging across both sides of the graft in absence of graft collapse and clinical evidence of healing"*. Eventhough no differences were reported, the external validity is limited by the missing information on observer (inter and intra-rater) reliability, precision of measurements, the X-ray set-up and software. The point is, that the broadly defined primary outcome measure and the insufficient description of the measurement method are issues encountered not only in Dolan's study but generally in the history of reporting results of paediatric CLOs. I will return to this in the chapter *"Implant durability - how do we know?"*

Hydroxyapatite tricalcium phosphate

Calcium orthophosphate bioceramics are promising alternatives to autologous bone graft. The great interest in calcium orthophosphate bioceramics can be ascertained to their biological safety, abscense of foreign body reaction, their osteoconductive capacity and their ability to facilitate osseous integration.[46,47] Since the commercialisation of calcium orthophosphate bioceramics for dental and surgical applications in 1980s, they have mainly been implemented as void fillers and coatings and to a lesser extent as structural bone graft substitutes.[48] Their porous and brittle nature has been the major limitation in their mechanical load bearing potential. Recently, research within bone tissue engineering has focused on improving the mechanical properties of calcium orthophosphate bioceramics,[43] in pursuit of an optimal and tailored composition of tricalciumphosphate (TCP) and hydroxyapatite (HA).

Figure II. Structural calcium orthophosphate bioceramics (ReproBone™).

Biphasic calcium bioceramics are composed of tricalcium phosphate, $Ca_3(PO_4)_2$, (β-TCP, or α-TCP) and hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$(HA).[49] HA is unique for its chemichal resemblence to mineralised bone, which serves as a moeity with osteoconductive potential and biocompatibility.[49] It is the crystalline form of TCP, hence HA dissolves slowly and provides mechanical stability to the implant. TCP is also bioabsorbable and biocompatible, but with a higher rate of degradation and less mechanical stability than HA. A combination of 60/40

HA/β-TCP yields a scaffold (Figure II), which facilitates osteoconductivity and osseous integration. Such a scaffold requires a porosity of at least 60 % (which is the percentage of void spaces in a solid including macroporosity and microporosity), and the pores should be highly interconnected. The suggested optimal macroporosity (pore diameter > 100 μm) should be in the range 200-400 μm.[49] This facilitates cell invasion from the host bone to the scaffold, leading to formation and nourishment of new bone and removal of waste products. Microporosity (pore diameter < 10 μm) should be at least 20 % of the total porosity.[49] Microporosity refers to the capacity of being impregnated by biological fluids. It provides a greater surface area for protein adsorption and greater ionic stability. An important property of the scaffold is the connection of pores. The higher interconnection, the more accessible the scaffold will be by gases, liquids and particulate suspensions all leading to cell adhesion, proper cell distribution, in vivo new blood vessel formation and thus, sustaining new bone tissue formation and remodelling.[48] The mechanical behaviour of HATCP is most often described by compressive strength of the material.[50] Cortical bone has a compressive strength in the range 90-230 MPa, whereas the compressive strength of cancellous bone ranges between 2-45 MPa.[50] The compresseive strength of porous HATCP resembles that of cancellous bone, thus applications of the material as a structural bone graft substitute in clincal settings have been sparse. Recently published experimental and clinical studies have shown good short-term results of the mechanical properties of HATCP. An experimental study on HATCP as an implant in K-wire-stabilised segmental tibial defect in rabbits with immediate postoperative weight-bearing revealed, that the implant did not fail at 18 weeks follow-up. Osseous integration was found in the most outerparts of the implant, mechanical stability was increased and the interconnected porous structure facilitated a favourable microenviroment for osseous integration.[51] The HATCP graft used in the study had a porosity of 63 %, macroporosity of 150-200 μm, microporosity of 1-2 μm and compression strength of 4.89±0.89 MPa.[51]

In a clinical prospective randomised trial by Gouin et al,[52] HATCP was compared with autologous graft (AUTO) in a valgus proximal opening wedge tibial osteotomy performed in 40 patients with a 2 year follow-up. Initially clinical and radiological union occured significantly later in the HATCP group than in the AUTO group, but with no significant difference six months after surgery. The implant used had a porosity of 75-80 %, porediameter of 300-500 μm and compressive strenght of 1-5 MPa.[52] Delecrin et al[53] investigated HATCP against autologous bone graft in a prospective randomised trial in 58 young adults, who underwent posterior spinal fusion. The patients were followed for approximately four years after surgery. During the follow-up period, correction was maintained clinically and radiologically (conventional X-rays) in both groups, with no significant differences.

To our knowledge, no randomised controlled study has yet been conducted on the clinical durability of structural HATCP compared with tricortical iliac crest autograft for CLOs in paediatric patients. In light of the recent studies proving no clinical significance between the HATCP and autologous bone graft, HATCP could potentially be noninferior in the ability to maintain lengthening of the osteotomy in a CLO of paediatric flexible PPV. Additional benfits such as reduced pain and elimination of donor site morbidy would be expected. There are three major reasons for why HATCP was considered as a potential structural bone graft: The required size of the trapzoidal-shaped bone graft is small, with the lateral side usually measuring between 1-1.5 cm. The foot is immobilised for at least six weeks in a below-knee cast with no weight-bearing allowed. This should be enough time to establish incorporation of bone within the implant without compromising structural durability. Finally, paediatric patients have a high bone healing potential.[54]

Implant durability

The ultimate goal with the CLO is to maintain the obtained correction over time and during functional loading. The patients should be asymptomatic. The implant provides a scaffold, which is later replaced by new bone from the host bed.[55] The new bone mass adapts and undergoes remodelling in response to functional loading. Clinically, the structural durability of the implant can be indirectly assessed by measuring the stability of the osteotomy.[56]

Until now, clinical studies in paediatric foot surgery have used conventional radiographs as primary outcome measurement instruments in the assessment of bone incorporation into the structural grafts.[16,19,41,44] From a clinical perspective this is a legitimate approach, as it reflects how bone incorporation is usually determined in daily practice. However, the generalisability of previous research on this topic is problematic, because of the vague or inconsistent descriptions of the criteria for bone incorporation, insufficient or absent descriptions of the applied X-ray set-up, no information on the analysis software and the precision of measurement methods.[16,44]

The assessments are typically based upon well-known radiographic signs such as trabecular continuity throughout the graft, loss of radio-opacity and discontinuity of the sharp margins of the graft. Height and length of graft can be measured using the distance measurement tool in the radiographic software available. Albeit clinically applicable, these parameters are observer-dependent. Anand et al[57] did an excellent reliability study on osteotomy healing defined as "callus bridging across three or four cortices at the osteotomy site" and found poor interrater and intrarater reliability and agreement. When comparing different bone grafts (or for that sake drugs, methods etc.), precision is a virtue, because the results have an impact on how future treatments are decided.

In lack of a method to measure the length of the calcaneus, Dayton et al[58] proposed a method to measure the mid-calcaneal length from weight-bearing lateral X-rays. The mid-calcaneal length was measured by drawing a line from the low point of the concave arc of the calcaneal-cuboid joint to the high point of the convex arc of the posterior facet. The method may seem appealing, but there are some shortcomings: It is an absolute measurement, thus external generalisation is limited if X-rays are calibrated differently. Growth potential in paediatric patients limits the use of this line for longer follow-up. If the foot position is not standardised (ie. the foot is angled differently), then the apparant distance might not be accurate.

Radiostereometric analysis (RSA)

Radiostereometric analysis (RSA), is a well-accepted method with high accuracy[59] for assessment of in vivo three-dimensional motion between bony structures (fracture healing, osteotomies, growth plate injuries, joint kinematics), implant fixation and wear.[60] With an accuracy between 0.2 mm and 0.3 mm for translations and 0.2°, and 1.2° for rotations in the assessment of prosthesis migrations it has become a valuable predictive tool in the phased introduction of new orthopaedic implants, primarily within hip and knee replacements.[61] In addition to the superior accuracy over conventional X-rays, the low radiation dose makes RSA a favourable tool for measurements of migration over time.

The use of RSA in children's lower extremities dates back to the early 1980's where it has been shown feasible to investigate growth patterns following various ankle injuries[62,63] and Crohn's disease[64] by inserting tantalum beads in the distal tibia and fibula. More recently Lauge-Pedersen et al.[65] used marker-based RSA to monitor percutaneous physiodesis within the distal femur and the proximal tibia, thereby providing information on the timing of physial arrest. The majority of the examinations have been performed with the patient positioned supine. The stereoradiographs can be obtained supine as well as standing, without affecting the clinical interpretation of the migration results. [66]

Feasibility and precision of marker-based RSA has been determined in smaller bones such as the metatarsal[67] and the trapeziometacarpal joint[68] in adults, and even to investigate joint kinematics in the talo-crural joint, by inserting tantalum beads in the distal tibia, fibula, talus, calcaneus, navicular, medial cuneiform and first metatarsal in healthy adults.[69]

Marker-based RSA relies on insertion of spherical tantalum beads into the region of interest (implant, bones). This creates well-defined measurement points which enable accurate measurements of migration from RSA radiographs. Tantalum is used owing to its several features, such as a high atomic number that facilitates identification on radiographs, sometimes even when shadowed by metal implants, its biocompatability and resistance to corrosion.[70,71] At least 3, but preferably more, non-colinear markers are combined in a group (rigid body) to define a 3D marker coordinate system. Appropriate and wide scattering of markers within the rigid body improve the accuracy of the RSA measurements.[72] The subject of interest is placed in front of/above a calibration cage and a pair of images are recorded by biplanar simultaneous radiographs (Figures III-IV). The calibration cage is a reference frame used to create and define the position and orientation of a global 3D coordinate system and to determine the position of the two radiographic foci.

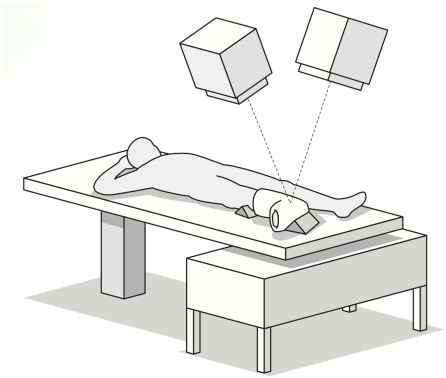


Figure III. RSA set-up with two radiographic tubes angled towards eachother. The patient's foot is positioned above a uniplanar calibration box.

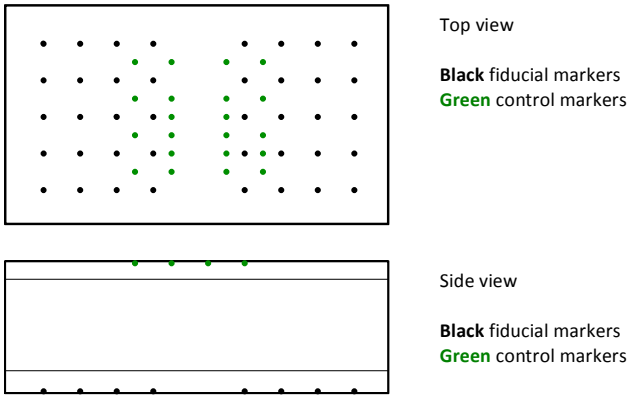


Figure IV. The calibration box viewed from the top and from the side with markers of known 3D postions. The plane nearest to the radiographic film contains the *fiducial* markers, which define the fiducial coordinate system. The superior plane contains the *control* markers, which are used to calculate the postions of the X-ray foci.

The images are processed by special-designed RSA software. Each of the marked rigid bodies is defined as a 3D model, and the absolute migrations (migration of a rigid body with reference to a fixed rigid body) are determined as translations and rotations within a Cartesian coordinate system. Considering the high accuracy and precision of RSA in measurements of three-dimensional absolute motion of rigid bodies, it would be a valuable tool for assessment of stability of the lengthening obtained with a structural graft material inserted the calcaneal bone.

The feasibility and precision of RSA measurements on CLO in paediatric patients have not been established before this thesis.

Patient reported outcome measures (PROMs) – which outcomes are meaningful to the patient?

In recent years, involving the patient's perspective on health-related quality of life has been gaining more attention in clinical trials,[73] as well as becoming an integrated part of clinical practice.[74] The method to collect this information in a standardised manner is by appropriate patient-reported outcome measures (PROMs). Initially PROMs were developed as a research tool to measure the patient's health. They were later embraced within various areas of clinical practice to enhance the patient treatment.[75] There are two main categories of PROMS: generic and disease-specific.[76] The generic types focus on quality of life in general, while the disease-specific are tailored to assess the symptoms and function related to a specific site/region or specific condition.

PROMs are considered valuable in clinical practice and research as an aid in the communication between healthcare provider and patient.[77] By elaborating on the patient's priorities, thereby defining the ultimate goals, which the patient and healthcare provider should strive to achieve, we can ensure an appropriate approach towards intervention and healthcare outcomes.[78,79]

Musculoskeletal disorders in the foot and ankle region comprise the majority of referrals of paediatric patients to specialised orthopaedic outpatient clinics.[2] Depending upon the extent to which the individual child is affected by his/her foot and ankle disorder, the optimal approach should embrace the wider impacts of the impairment on the patient - as have been well defined in the framework of WHO's International Classification of Functioning, Disability And Health (ICF).[80] This model integrates the medical and social model of disability. Disability and human functioning are considered as a result of the interactions between the "body functions and structure", the "whole person", and how the whole person functions within a social context. Thus, disability becomes a function of the medical and social variables in a human being. Therefore, both aspects must be considered in relation to choosing the appropriate intervention.

Current practice include anamnesis, clinical examination and para-clinical measurements such as various imaging techniques, e.g. pedobarography, gait analysis, x-rays. These measurements do to an extent quantify the physical manifestations of the subject matter, but how about the impact on participation in daily activities and personal and enviromental aspects – do we always get the full picture? Such considerations have led to the developement of the disease-specific PROM, the Oxford Ankle Foot Questionnaire (OxAFQ).[81] The OxAFQ was presented in 2008 to assess disability associated with foot and ankle issues in children aged from 5 to 16 years, as reported by the child and/or parent. Comprising a total of 15 items, divided into three domains; Physical activity (six items), School and Play (four items), Emotional (four items) including an additional single footwear item, the OxAFQ was meant to assess the well-being of the child within an orthopaedic framework. The OxAFQ has been evaluated by the developers with focus on the

discriminative properties, test-retest reliability, child-parent reliability, internal consistency, construct-, content- and longitudinal validity, in children with various foot and ankle disorders.[81,82] The established measurement properties of the OxAFQ as well as the cultural resemblance make it an attractive PROM to translate and culturally adapt to paediatric's orthopaedics in Denmark. The surgeon could be equipped with a valid, reliable and feasible instrument to aid in the "patient-doctor" communication in the assessment of the impact of foot and ankle disorders on children's lives.

Prior to implementing any translated PROM within the research or the clinical setting, the measurement properties of the PROM should be evaluated in the new language; this is termed cross-cultural validation.[83] Obscure definitions of what constitutes a good measurement property and how to measure it led to the foundation of the COSMIN initiative where a set of tools were developed for design, definitions and statistical methods to ensure methodological quality of studies on the subject matter. In this study the taxonomy provided by the COSMIN[84] was chosen.

The COSMIN initiative describes following taxonomy for PROMs which are to be used as an evalutative tool: Reliability, validity, responsiveness and interpretability. Responsiveness is not within the scope of this study, but we have an ongoing study designed to evaluate this domain.

Reliability

Internal consistency: The extent to which item scores correlate with each other in a given domain. Can be assessed by Crohnbach's alpha.[85]
Reliability (Test-retest and inter-rater reliability). The relative measure of reliability (reflects consistency of scores). Refers to the proportion of the total variance in the measurements, which is due to the true difference between patients. "True" is the average score, that would be obtained if the scale was given an infinite number of times. Can be assessed by the Intraclass correlation coefficient (ICC).[86]
Measurement error (Test-retest and inter-rater). The absolute measure of reliability. Represents the systematic and the random variation in scores not attributed to true changes. Can be assessed by Bland-Altman statistics.[87]

Validity

Construct validity (hypothesis testing). The degree to which the scores of the measurement instrument are consistent with predefined hypothesis based on the assumption that the intstrument validly measures the construct to be measured. Hypothesis testing will be elucidated under the Method section.

Interpretability (feasibility)

Number of completed items, time to completion, floor and ceiling effects. Floor and ceiling effects refer to the fraction of the respondents who reported the lowest or highest possible scores. Limits ranging from 15 to 30 % have been used.[88,89] If these are present, it may indicate a limited content validity in the patients who are in the lower or in the higher end of these scales. Consequently, this may impair the interpretation of these patient scores. In this thesis, a limit of 30% was chosen, as it was considered appropriate for this study population of foot and ankle patients, as the emotional domain in the OxAFQ was not expected to be severely affected.

Hypotheses

Does HATCP graft provide acceptable structural durability as compared to tricortical iliac crest autograft in patients who undergo calcaneal lengthening osteotomies for symptomatic flexible PPV?

RSA is used to measure the primary outcome, which is the structural durability of the bone graft materials, and the OxAFQ is used as the secondary outcome measurement instrument to describe the patient's perspective on the results. To ensure validity of the results, precision and feasibility of these methods must be determined. Each of the three studies in this thesis seeks to contribute an answer to the overall research question. *Study I* seeks to translate the English version of OxAFQ into Danish and secondly to validated the translated OxAFQ (child and parent versions). *Study II* seeks to investigate the precision and feasibilty of RSA measurements of CLO in cadaver feet. *Study III* seeks to investigate noninferiority of HATCP to tricortical iliac crest autograft (AUTO) as a structural bone graft for CLO in paediatric flexible flatfeet. Interim results are presented after inclusion of half of the final sample size at six months follow-up. The primary outcome is osteotomy stability (x-translations) measured by RSA.

Hypothesis I

The Danish translation of the OxAFQ is valid in a paediatric population encountered in the outpatient clinics of children's orthopaedics.

Hypothesis II

RSA measurements of CLO in cadaver feet are precise and feasibile.

Hypothesis III

HATCP is noninferior to AUTO graft in maintaining stability of the obtained lengthening of the calcaneus in calcaneal lengthening osteotomies for paediatric flexible flatfeet when measured with RSA six months after surgery.

Design

Study I

Cross-sectional study.
Translation performed in accordance with the ISPOR Task Force Translation guidelines. [83]
Validation performed in accordance with the COMSIN Guidelines[84]

Study II

Experimental study.
Conducted and reported according to the STARD criteria.[90]

Study III

Prospective, randomised, controlled noninferiority study.
Conducted and reported according to the CONSORT noninferiority extension guidelines.[91]

Materials & methods

Ethics

Study III was conducted in accordance with the Helsinki declaration of 1995. The original protocol and a supplementary version were both reviewed and approved by the Central Denmark Region Committées on Biomedical Research Ethics (Record No. 1-10-72-250-12, Appendix No. 40345 Issue dates 24.04.2012 and 14.11.2013). The protocol was registered in a public clinical trial database (ClinicalTrials.gov NCT01770574). *Study I* and *II* did not require ethical approval. The Danish Data Protection Agency was notified on all the projects (Jr. No. 1-16-02-86-12).

Patients

Study I

It was a prerequisite that the participants possessed good Danish language skills. Participants were considered eligible for inclusion if they met the following criteria:

- Aged 5 to 16 years
- Accompanied by a parent or a guardian
- Referred to the Children's Orthopaedics outpatient clinic at Aarhus University Hospital for a foot or/and an ankle disorder without any significant proximal musculoskeletal component.

Study II

This study comprised three fixed cadaveric human species (six feet), all male of Caucasian origin. Mean age was 77 years (range 73-88), the feet measured a European size 40 and there was no prior history of foot surgery. Death causes were prostate cancer (2) and unknown (1).

The fixing liquid compound contained 5 L alchohol 96 %, 2.5 L glycerine, 2 L destillated water and 1 L formaline. The cadavers were stored at 6 °C and each body was tempered for a minimum of 60 minutes prior to the operation in an air conditioned room with a temperature of 22 °C. The surgeries and RSA recordings were performed within one day at Aarhus University Hospital (AUH).

Study III

Patients were eligible for participation if they had been scheduled for elective CLO surgery based on the indication of symptomatic, flexible PPV at AUH. They patients had to be between 5-16 years, with intact ambulatory function, no major cognitive deficits and no planned simultaneous major ipsilateral orthopaedic surgical procedures.

The flatfoot was a clinical diagnosis made by the operating surgeon. It was defined by collapse of the medial longitudinal arch during stance, hindfoot valgus and abduction or neutral position of the forefoot. Subtalar joint mobility was present if the calcaneus inverted during the double-heel raise test. Tightness of the Achilles complex was assessed by the Silfverskiöld test.[92]

Patients would be excluded from the study in case of major complications that would require removal of the graft material.

Interventions

Translation and validation of the OxAFQ

The first objective of *Study I* was to translate the OxAFQ in accordance to Wild et al.[83] In preparation to this proces, permission was obtained by the copyright holders, Isis Innovation Ltd.

The *forward translation* from English to Danish was performed independently by two medical doctors, who were both native speakers of the target language, (Danish), residing in the target country (Denmark) and also fluent in the source language (English). The two forward translations were sent to the in-country investigator, who produced a *reconciled version* of the translation. The reconciled version was *translated back* into English independently by two translators, who were blinded to the original English version. Finally three investigators performed a *comprehensive review* of the back translation against the original version in order to make revisions and eliminate discrepancies. The Danish version of the OxAFQ (comprising both the child and parent versions throughout this dissertation) was pilot tested within ten heterogeneous Danish child-parent pairs in an orthopaedic outpatient clinic setting. Participants puzzled over item 14, but found it comprehensible. Few terms were changed to concur with the Danish culture. The copyright holder reviewed and accepted the final changes. *Proofreading* was performed by the in-country investigator and another translator, who had not been involved in the prior steps.

The second objective of the study was to validate the Danish version of the OxAFQ (cross-cultural validation). The validation proces was conducted in accordance with the taxonomy proposed by COSMIN guidelines.[84] The questionnaire was given to a subsequent study population (84 children and 83 parents of which 82 children and 81 parent participated) on two occasions at two different time points (1-2 days) in order to assess reliability and measurement error (test-retest and inter-rater).

The initial test. Eligble patients and their parents were approached in the Children's Orthopaedics Outpatient clinic by the principal investigator or an assistant, prior to their scheduled consultation. Each patient-parent pair received the OxAFQ and the CHQ[93] (child version, long, CHQ-CF87 and parent version, short, CHQ-PF28). Parents were allowed to read out the questionnaire for their children, but were advised to refrain from providing any interpretations/explanations. For children below nine years, only the parents were requested to complete the proxy version. This had been recommended by the Danish validator of the CHQ, who recognised interpretability issues in children below this age (advise obtained by e-mail correspondence). Overall participants were divided into three groups: Those who

were newly referred to the orthopaedic clinic (group 0), those who were managed conservatively in the outpatient clinic (group 1) and those who were scheduled for elective surgery (group 2).

The retest. After completion of the first set of questionnaires, the participants received a prepaid envelope containing another set of questionnaires, which they were requested to complete and return by mail after two days. Partipants received a reminder text-message on the retest date.

Operative procedure and bead marking

Study II and III

The CLO followed the standard procedure of the institution. A lateral longitudinal skin incision was initiated proximally to the calcaneal-cuboid joint (CCJ), expanding distally and plantarly, and further over the cuboid bone to gain access for subsequent bone markings. The sural nerve was spared. The extensor digitorum brevis muscle was incised proximally and retracted from the dorsal aspect of the calcaneus together with the soft-tissue of the sinus tarsi. The peroneal tendons were released from their tendon sheet and retracted plantarly. The calcaneal bone was exposed and a Joker periostal elevator was used to probe the dorsal and plantar aspect of calcaneus on the most narrow part of the bone. The periosteum was incised at the osteotomy site, about one cm proximally and parallel to the CCJ. The osteotomy was done with an oscillating saw parallel to the CCJ. Distraction was done and maintained with a laminar spreader.

In preperation for radiostereometry, the distal and proximal fragments of the calcaneus (dCF and pCF) and the cuboid were each marked with four tantalum beads (0.8 mm beads used for dCF and 1.0 mm beads used for pCF and the cuboid), which were as widely distributed as possible. The markers were inserted by bead-injectors (Kulkanon, Wennbergs Finmek AB, Sweden) (Figure V), and the marker-position was controlled by fluoroscopy. The dCF and pCF were accessed through the osteotomy. The cuboid was marked to investigate the CCJ for a different study, which will not be discussed further in this thesis.

In the *Study II* the cuboid bone was also marked with 0.5 mm beads. A customised trapezoidal-shaped plexi-glass wedge was used for distraction. The lateral border of the wedge was marked at 1 cm, 1.25 cm and 1.5 cm, and the osteotomy distraction was done corresponding to each marking by pushing in the wedge.

In *Study III* the osteotomy was dilated until the clinical medial arch was restored on simulated weight-bearing, judged by the surgeon. If the patient was allocated to autograft, iliac crest bone graft harvesting was approached by a 3 cm antero-lateral incision, 1 cm

below the iliac crest. Soft tissues were retracted and the apophysis was split and iliac bone was exposed subperiostally. Tricortical trapezoidal-shaped bone grafts were harvested with an osteotome, with the length of the base matching the required length of correction. Following graft harvesting, the donor site was packed with a haemostatic gelatinous sponge (Spongostan™) and sutured with Vicryl® 3.0 mm. When the HATCP bone graft substitue was used, it was often necessary to trim the graft after insertion because some graft protruded laterally. Trimming was done with a knife, but debris distribution of the graft material into the surrounding tissues was difficult to avoid, even with repeated wash-outs.

Tightness of the Achilles tendon or gastrocnemii muscles was assessed by the Silfverskiöld test[92] and adressed by percutaneous lengthening when necessary.
The foot was immobilised with the ankle in neutral in a below knee X-lite circular cast. No weight bearing was allowed for six weeks.

The preoperative administration of antibiotics followed a standard regime with either dicloxacilline or cefuroxime as single dose intravenous injections. The surgery was performed under general anaesthesia. The standard analgesia targeted at the iliac crest was a subcutaneous injection of about 10 mL ropivacaine (concentration 7.5 mg/mL). The osteotomy site received a nerve block (popliteus/sapheneous/tibialis) including a single shot and a catheter with continuous infusion of body-weight adjusted (~30mg/kg) ropivacaine (2mg/mL), rate 3-5 mL/hour, discontinued after 48 hours. The standard postoperative analgesia regime of the department was paracetamol 500 mg x 4 per 24 hrs with a combination of long-acting and short-acting opioid equivalents, adjusted on an individual basis.

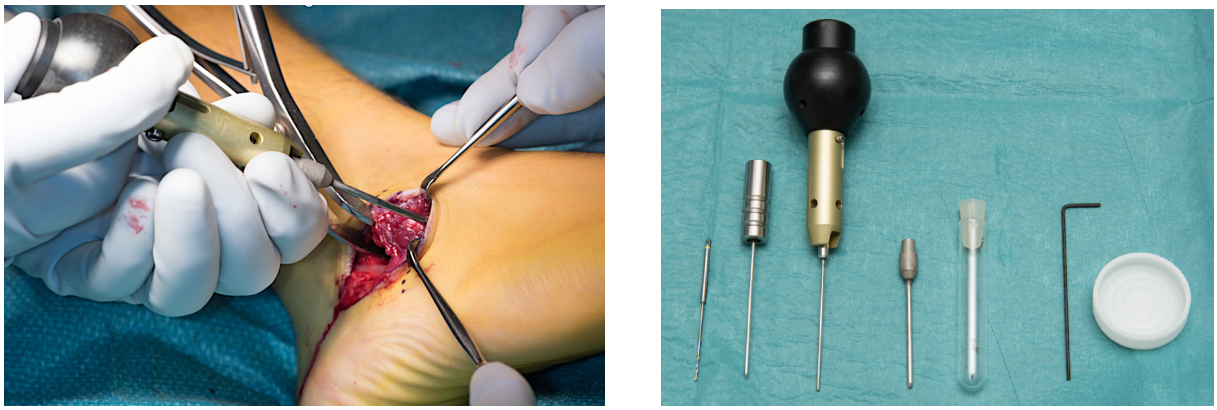


Figure V. Bead injection in the cuboid (*left*) and instruments for bead insertion (*right*)

Hydroxyapatite tricalcium phosphate

The HATCP material (ReproBone™) was provided by Ceramisys Ltd© (Sheffield, England). It is a biphasic calcium-ceramic with a 60/40 ratio composition of hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$, HA and β-tricalciumphosphate (β-$Ca_3(PO4)_2$), β-TCP. It comes as a trapezoidal-shaped block, available in different sizes. The HATCP has fully interconnected pores, the porosity is 83 %; macroporosity between 250-800 μm and a microporosity between 1.0-10 μm. The sintering temperature is higher than 1200 °C and the compressive strength ranges between 4-7 MPa. The information is obtained from Ceramisys Ltd©.

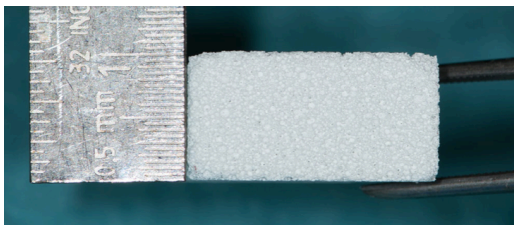


Figure VI. HATCP graft. The lateral side required for lengthening is measured.

Figure VII. Iliac crest bone grafts.

Primary outcomes

Validity parameters for the OxAFQ

In *Study I* The Danish version of the OxAFQ underwent cross-cultural validation. With reference to the COSMIN guidelines this included reliability (internal consistency, test-retest reliability, child-parent reliability), validity (construct validity by hypothesis testing) and interpretability (number of completed items, time to completion, floor and celing effects). The test-retest reliability was assessed in the group who completed the retest. We were interested to know if the OxAFQ was able to discrimminate between patient groups, primarily within the physcial domain. We expected that the patients scheduled for elective surgery would have lower physical scores than those attending clinical control (Hypothesis Number I, t-test). Another eight hypotheses on concurrent (and convergent) validity were constructed by correlating OxAFQ domains with similar domains of the CHQ. Spearmans rho, correlation of ranks was used as statistical parameter.

The following eight hypotheses (No. II-IX) were expected to correlate with each other.

II. Physcial domain and Physical Functioning
III. Physical domain and Role/Social Limitations - physical
IV. Physical domain and Bodily Pain
V. School & Play domain and Physical Functioning
VI. School & Play domain and Role/Social Limitations - physical
VII. School & Play domain and Bodily Pain
VIII. Emotional domain and Mental Health
IX. Emotional domain and Self-Esteem

Radiostereometric analysis

RSA was the primary outcome measurement instrument in *Studies II and III*. X,y,z-translations of the osteotomy were the primary outcomes in *Study II* and x-translations of the osteotomy was the primary outcome measure in *Study III*.

The RSA set-up is reported in a standardised manner with reference to the description in *Paper II and Paper III*.[94]
"A standard set-up of two synchronised ceiling fixed X-ray tubes (Arco-Ceil/Medira; Santax Medico, Aarhus, Denmark) angled towards each other at 40° and a focus-grid uniplanar carbon calibration box (Box 24, Medis Special, Leiden, The Netherlands) was used (Figure VIII). The calibration box defined the position and orientation of the global coordinate system. The exposure was set to 70 kV, 10 mAs, fine focus." Initially, and during the cadaver study (II) all radiographs were digital computed radiography (FCR Profect CS, Fujifilm, Tvedbæk, Denmark) with a resolution of 2.5 lp/mm. After one year in clinical the project (III), the RSA system was upgraded with an AdoraRSA suite (Nordic Roentgen Technique, Hasselager, Denmark) using high quality DR (digital radiography) imaging technology with a resolution of 4 lp/mm (Canon CXDI 50RF detectors for supine positions and 70C detectors for standing positions).

Figure VIII. Supine RSA set-up (*left*). The patient was positioned supine with the operated foot aligned parallel to the uniplanar calibration cage, the lateral side of the foot facing the calibration cage, the medial side facing upwards towards the detectors and the ankle in neutral position. The focus was centred on the CCJ. The cast was used as a template for all subsequent follow-ups. Standing RSA set-up (*right*) matching the supine set-up in tube-angulation and source-image-distance (SID) but the foot is weight-bearing and non-casted. (Images are used with the permission of the parent, children and radiographer).

In *Study II* double stereoradiographs were obtained with the osteotomy in zero distraction and at described markings (1 cm, 1.25 cm and 1.5 cm). The foot was repositioned within each double-examination.

In *Study III* supine stereoradiographs were obtained within two days after surgery (baseline), at six weeks, eight weeks, six months and 12 months. Double-examinations were scheduled at six weeks follow-up, to assess the clinical precision/repeatability of the RSA measurements. Patients were repositioned within each double-examination with the operated foot in cast. The cast was used for subsequent follow-ups to ensure consistency in the foot position.

Bone marker models and analysis

Marker-based analysis of all stereoradiographs was performed by the primary investigaor, using Model-based RSA 3.32 software (RSA*core*, Leiden, The Netherlands). To ensure consistency in migration data, the analysis criteria requires the condition number (CN) and marker number within each individual bone marker model to be consistent throughout all scenes, and a rigid body error (RBE) threshold of maximum 0.5 mm.[95] Bone markers with a RBE exceeding 0.5 mm are automatially detected and excluded by the software, as they are by default considered unstable. The upper limit for CN was set to 350 (and 300 in *Study II*) which was of higher limit than recommended by Valster et al,[95] but was a natural result of the small bone size of the marked anatomical regions. The CN reflects the spatial dispersion of the markers within a bone fragment. Geometrically, it can be interpreted as the square root of the inverse of the square sum of the distances of markers to a straight line.[96] The smaller the CN, the further the markers are distributed away from each other and from a straight line. Correct marker matching is crucial for calculating micromotion between rigid bodies (bone models) over time, which is why analysis criteria require consistency of CN and marker number within each individual. At least 4 non-co-linear markers should be inserted in each bone region of interest in case marker loosening should occur.

Three bone marker models were created (Figures IX-X): The distal/anterior calcaneal fragment (dCF), proximal/posterior calcaneal fragment (pCF) and the cuboid. The osteotomy migrations were described with the reference model being the pCF and the migrating model being the dCF. The CCJ migrations were described by the cuboid as the migration object with reference to the dCF. Figure IX illustrates the marker-based analysis and Figure X describes the translations for the osteotomy.

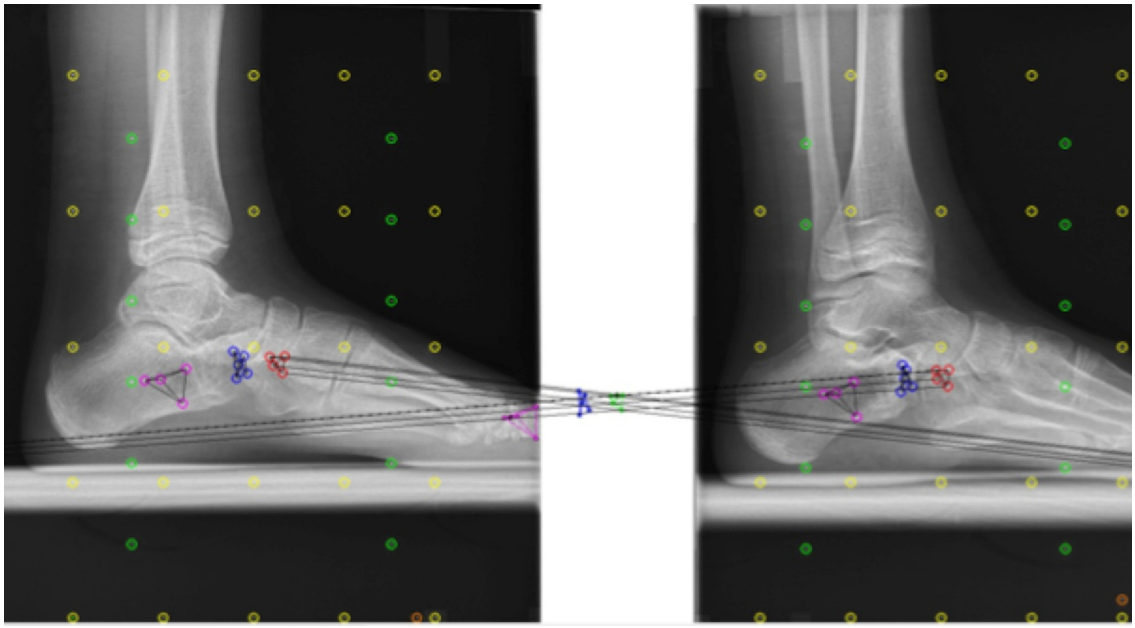


Figure IX. Marker-based analysis of the osteotomy (the analysis of the CCJ is not within the scope of this thesis). Standing RSA example provided with focus on CCJ migrations. Three bone models are depicted: pCF (magenta), dCF (blue) and cuboid (red). The bone models can easily be distinguished from each other.

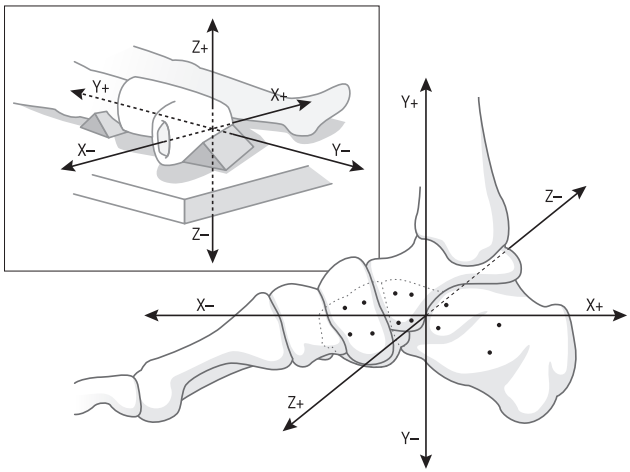


Figure X. Adapted from *Paper II and Paper III with permission*. The positioning of the foot during RSA examinations. The foot is viewed from the medial side. There are four markers in the cuboid, the dCF and the pCF. The right handed coordinate system with signs represents the direction of translations. Interpretation of migration in the CLO is described by the migration of the dCF along and about the 3 orthogonal axes with reference to the marker-model of the pCF as follows: x-translation (+osteotomy distraction/-osteotomy compression), y-translation (+dorsal migration/-plantar migration), z-translation (+medial/-lateral), x-rotation (+medial rotation/-lateral rotation), y-rotation (+internal rotation/-external rotation), z-rotation (+plantar rotation/-dorsal rotation).

Secondary outcome measures

(Only relevant for study III)

Health related quality of life - measured by OxAFQ

The validated Danish version of the OxAFQ was used as a secondary outcome measure to assess how the surgical intervention affected physical function, participation in school and play, and footwear. The impact on emotional well-being was not expected to be significant, as the surgical intervention is principally aimed towards improving the physical function. Patients and their parents each completed a questionnaire at the preassessment date, at six and 12 months follow-up. Parents served as a proxy for children who were not able to complete the questionnaire.

Complications

Reported according to the new and reliable complication classification system by Dindo et al.[97]

Pain

Patients were scored before surgery and postoperatively at 3 hrs, 5 hrs, 8 hrs, 12 hrs, 24 hrs and at discharge. A pain assessment sheet was created specifically for the project containing the numerical rang scale (NRS), visual analogue scale (VAS), and faces pain scale by Wong-Baker and Brieris (Appendix 2). The primary parameter of interest was the NRS, which is routinely used at the ward and is also validated for measurement of acute pain in children after major surgery.[98] Children were instructed to complete the pain ratings by considering how much overall pain they had, and subsequently to point out the region where the greatest pain was experienced. The time points were approximate and adjusted on individual basis. Patients were not awakened for pain scoring. Supplementary analgesics were given either prior to pain assessment, or the pain assessment was performed at least one hour after administration of analgesics. Patients who received the AUTO graft were asked about their hip pain (presence/absence) at each follow-up time (6 weeks, 8 weeks, 3 months (at the surgeon's discretion), 6 months and 12 months).

Analgesics

Consumption of all analgesics after surgery was registered in the Electronic Patient Journal (EPJ) chart.
Duration of surgery (knife time) and length of hospital stay were registered in EPJ as well.

Radiographs

Obtained in two projections (supine: lateral and oblique, standing: antero-posterior and lateral). Before surgery, standing or supine radiographic examination were performed to confirm the bony malalignments. At six weeks after surgery supine radiographs were obtained by the radiogoly department at Nørrebrogade, AUH to assess signs of bone incorporation into the graft prior to cast removal. This included a qualitative assessment of trabecular continuity, loss of radiopacity and gradual blurring of the sharp margins.

Arch index

Patients were enrolled in a dynamic plantar pressure protocol (Appendix 1) with examinations scheduled at the preassessment, six and 12 months follow-up. In this dissertation, the analysis is confined to a calculation of the arch index (AI) to provide baseline patient characteristics.
We adapted the definition by Cavanagh et al.[99]

Arch index = $CA_{midfoot} / (CA_{midfoot}+CA_{hindfoot}+CA_{forefoot})$ *(not including the toes).*

Statistical Methods

In *Study I* the sample size in this validation is based on the intended assessment of measurement error by using Bland-Altman statistics. With reference to suggestions by Bland-Altman[100] we accepted an accuracy of 0.4 x SD, thus requiring approximately 70 participants. Hypothesis testing was approached by t-tests (group comparison) and Spearmans rho (comparison with CHQ; non-parametric correlation was suggested by our statistician). Internal consistency was approached by calculating Crohnbach's alpha.[85]

In *Study II and Study III* repeatability/precision of the RSA system was assessed in accordance with the International Standard and definitions (ISO 1998) by double-examinations. The first examination served as a reference scene. The systematic variation (bias) was expressed as the mean difference, $mean_{diff}$, between the first and second part of a double-examination. The random variation (variance) was calculated as the standard deviation of differences, SD_{diff}, between the first and the second measurement in a double-examination along with the repeatability limit, expressed as coefficient of repeability: $\pm 1.96 * SD_{diff}$, which represents the lower limit of individual actual migrations.[95,101] In study II the precision was expressed as the 95% limit of agreement, LOA, representing the expected clinical precision.
The sample size of *Study II* is based on former research on RSA examinations of the trapezometacarpal joint.[102] It was anticipated that six cadaver feet would be sufficient to investigate both feasibility and precision of RSA measurements on CLO.

Interim guidelines and stopping guidelines

Sample size

The design of *Study III* warrants a seperate description of statistical methods.
Sample size calculations followed the noninferiority continuous outcome trial method, developed by Sealed Envelope Ltd. 2012.[103] Noninferiority of the HATCP graft would be claimed if the upper limit of a one-sided 95% confidence interval (CI) with a significance level of 5%, a power of 90%, a standard deviation of 1.5 mm (both groups) would be below a predefined noninferiority limit (Δ_{NI}) of 2 mm x-translation.[1] This would require a sample size of 10 patients in each group. To ensure enough data in case of drop-out, 30 randomisation envelopes were generated. The Δ_{NI} was based upon the senior surgeon's clinical experience of a "minimally clinically important difference". It represents an estimate of how much more

[1] (+) x-translation: osteotomy "compression" and (-) x-translation: osteotomy "distraction".

loss of lengthening could be accepted by using HATCP compared with AUTO bone graft. Hence, the primary efficacy endpoint was the mean change in x-translation (Tx) of the distal dCF with reference to the pCF at 6 months after surgery, and is thus a surrogate measure of graft compression.

Interim analysis

Paper III which is the base for *Study III* is based on an interim analysis, as defined by reaching ten patients (halv of the final sample size) with six months follow-up. The primary outcome was analysed with relation to the Δ_{NI}. Non-inferiority was claimed if the upper limit of a two-sided 99.8% CI would be below the Δ_{NI} of 2 mm (new treatment minus old treatment).

Stopping guidelines

For safety reasons, if the new treatment was suspected to be worse, stopping guidelines were predefined. We would consider the difference in x-translation from a two-sided 99.8% CI at 6 months between the two groups. If the lower limit of the 99.8% CI was above zero while the upper limit above 2 mm, or even worse if both limits were above 2 mm, then the study would be stopped in fear of harmful effect. Possible scenarios anticipated for this study are described in Appendix 3. Should adverse effects occur, that would require surgical intervention (grade III complications[97]) related to the new bone graft substitute, the study would be stopped.

Sensitivity analysis

In genereal, study data were analysed by a per-protocol (PP) principle. However, in case of loss of follow-up both PP and intention-to-treat (ITT) principles were applied to assess the robustness of the primary outcome measure (the x-translation within the osteotomy).

Statistical software

Statistical analyses were performed in Excel version 14.3.5 (Microsoft Excel for Mac 2011, Redmond, Washington) Stata IC 12.1 for Mac (StataCorp, Texas, USA). Normality was assessed by QQ-plots and histograms (*Study I and II*) or by Shapiro-Wilk test (*Study III*). Equal variance was checked by f-test. The level of statistical significance was set at $p<0.05$. Unpaired t-tests were used for group comparisons.

Summary of Results

Patient characteristics

In *Study I* eighty-four child-parent pairs were invited to participate, of which two children and three parents declined providing no particular reason. One of the parents to the included child was a father, who did not see his child often, and thus did not feel he was capable of answering the questionnaires. A description of the patient characteristics are given in Table I and baseline measures at the first and second test are reported in Table II. Physical function and the foot wear item appear to be the most impaired domains in relation foot/ankle deformities.

Table I. Demographics (mean, range) and diagnosis for n=82 children

1st Test	Child	Parent
Subjects (n)	82	81
Age (years)	10.7 (5.5-16)	NA
Male/female	36/46	15/66
2nd Test (retest)	Child	Parent
Subjects (n)	54	49.4 (33-65)
Age (years)	11.5(5.5-16)	NA
Male/female	24/30	15/38
Diagnosis (total n = 85[a])		
Pes planovalgus/pes planus	22	
Congenital talipes equinovarus	10	
Idiopathic toewalker	8	
Apophysitis/tendinitis	8	
Macrodactylia	5	
Minor deformitites[b]	12	
Trauma[c]	7	
Other[d]	13	

[a]Some had more than one diagnosis
[b]Curly toes, hallux valgus, metatarsal deformities, os naviculare accessorium, syndactyly
[c]Trauma sequelae
[d]Multiple hereditary exostoses, ingrown infected nails, osteoid osteoma, restricted ankle range of movement, drop foot, short lateral malleolus, tarsal coalition and non-specified pain

Table II. Mean(SD) and range for baseline and retest scores (%) in children and parents. Adapted from *Paper I*.

	Baseline			Retest		
	n	Mean(SD)	Range	n	Mean(SD)	Range
Physical						
Child	82	62.5(22.5)	12.5-100	54	67.5(22.3)	7.5-100
Parent	81	59.5(24.3)	0-100	53	67.5(22.3)	20.8-100
School & Play						
Child	82	77.5(23.4)	6.25-100	54	80(22.5)	25-100
Parent	81	77(23.3)	18.8-100	53	82.5(21.2)	25-100
Emotional						
Child	82	82.5(17.5)	6.25-100	54	85(18.8)	25-100
Parent	81	78.3(20)	12.5-100	53	85(14)	43.4-100
Footwear item						
Child	82	57.5(37)	0-100	54	57.5 (32.5)	0-100
Parent	81	52.5(35)	0-100	53	57.5(37.5)	0-100

The patient flow in *Study III* is described in Figure XI.

23 patients were found eligble for inclusion and two patients did not meet the inclusion criteria, and where thus not presented for the study. Of the 23 patients, nine declined to participate. Various reasons for refusing to participate were given: Parents uncomfortable with not being able to make the decision of which treatment their child should have. Skepticism towards articifial bone graft material. Why change a method that is already good? Refusal of attending the hospital more than necessary. Three patients were not invited to participate.

Baseline characteristics of the patients are described in Table III.

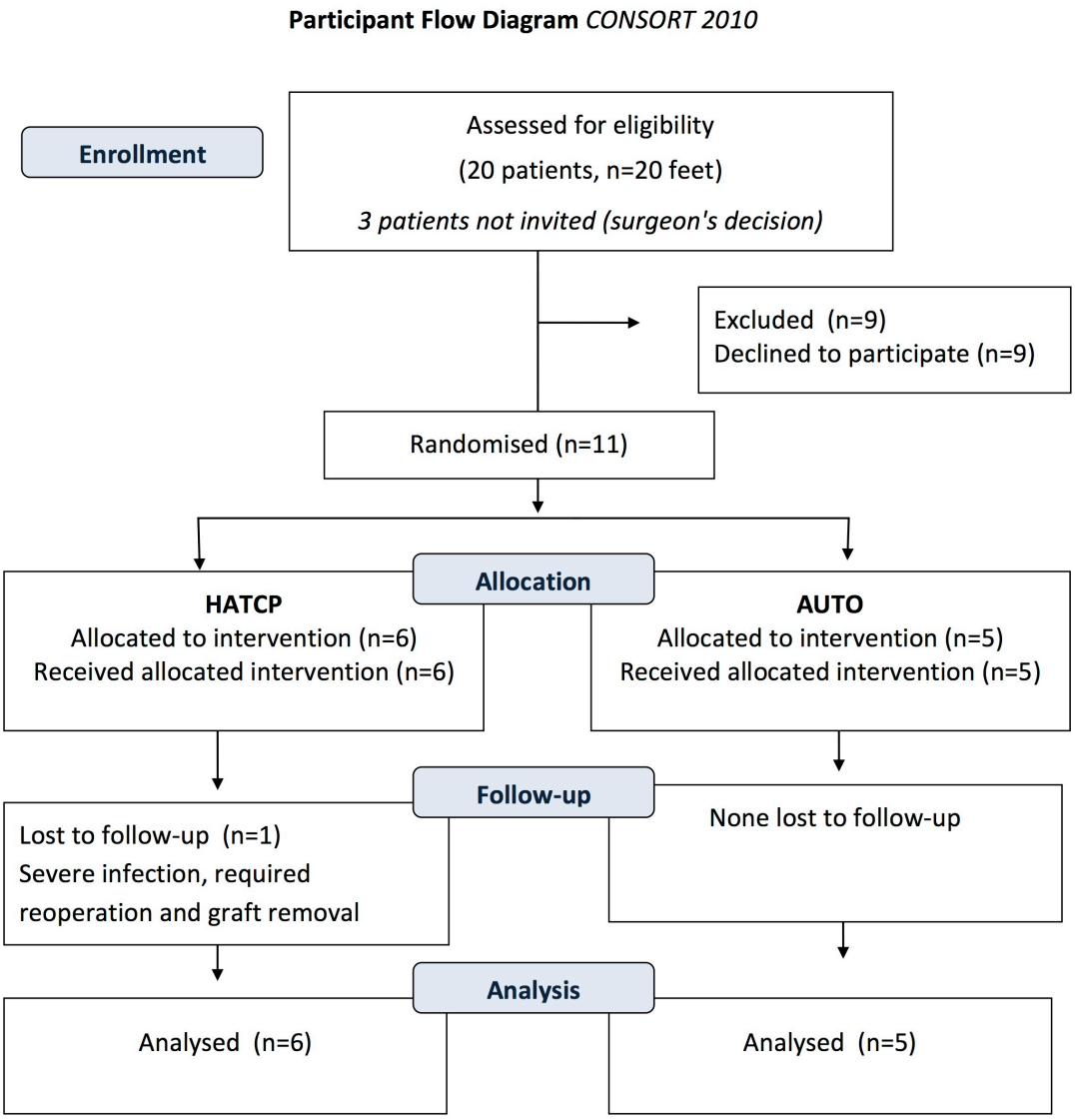


Figure XI. Participant flow in *Study III*. Adapted from *Paper III*.

Table III. Adapted from *Paper II*. Demographics, reported as mean and range (n=11).

Patient variables	AUTO group	HATCP group
Subjects (n)	5	6
Age (years)	10.8 (8.2-13.7)	12.3 (9.9-14.2)
Male/female	2/3	1/5
Height (cm)	144.4 (137-162)	156.6 (132-164)
Weight (kg)	35.18 (28-47.7)	49.4 (33-65)
Right/left	2/3	2/4
PETA	3	1
[a]Arch index	0.3(0.1-0.4)	0.3(0.1-0.4)
Preop X-rays	AUTO (n=4)	HATCP(n=3)
[b]Lat Talo-1M (°)	10 (7-12)	14 (9-22)
[c]Lat CP(°)	14 (10-20)	15 (13-18)
[d]Lat TC(°)	43(29-50)	42 (33-52)

Other diagnoses: Idopathic juvenile psoriasis arthritis (IJPA, 1), mild psychomotorical retardation (2), cerebral palsy (hemiplegia, 2) and concomitant contralateral eight plate removal and revision of iliac crest scar from former surgery (1).

[a]Arch index = $CA_{midfoot}/(CA_{midfoot}+CA_{hindfoot}+CA_{forefoot})$[99]

[b]Lateral talo-1st-metatarsal angle, normal 13°±8[104]

[c]Lateral calcaneal pitch, normal 17°±6[104]

[d]Talo-calcaneal angle, normal 49°±7[104]

Outcome measures for study I

Reliability, internal consistency and validity

Table IV. Description of relative (ICC) and absolute measurement error (%) for children and parents responses. Summary of tables adapted from *Paper I* Martinkevich et al[105]

Test-retest (intra-rater) reliability			
Domains	**n**	**Mean$_{diff}$± SD$_{diff}$**	**ICC**
Physical			
Child	54	-3±14.3	0.8
Parent	53	-3.3±14	0.8
School & Play			
Child	54	0.1±11.3	0.8
Parent	53	-3±11.4	0.9
Emotional			
Child	54	0.1±15.6	0.6
Parent	53	-2.3±12.6	0.8
Footwear item			
Child	54	3.4±23	0.8
Parent	53	1.5±26.1	0.7
Child-parent (inter-rater) reliability			
Domains	**n**	**Mean$_{diff}$± SD$_{diff}$**	**ICC**
Physical	81	3±19.3	0.68
School & Play	81	1.5±18.8	0.69
Emotional	81	5± 18.7	0.51
Footwear item	81	5.3± 37.8	0.49

Internal consistency, expressed as Crohnbach's alpha ranged from 0.75 to 0.88 for parents and from 0.67 to 0.87 for children, reflecting an overall good internal consistency in the item responses. Test-restest and child-parent reliability was good for the domains, but with a larger variation in the foot wear item.

The construct validity is expressed in Table V.

Table V. Hypothesis testing. No. I: Group 2 has lower physical scores than group 1 (t-test). No. II-IX OxAFQ domains correlated with CHQ (CF87 and PF28).

		Child		Parent	
No.	**Hypothesis**	**$Mean_{diff}$(95% CI)**	**p-value**	**$Mean_{diff}$(95% CI)**	**p-value**
I	Elective (n=11) < clinical FU (n=22) Physical scores	38.8 (23.8; 55)	0.0002	18(2.3; 33.7)	0.0001
		Rho (n=65)	p-value	Rho (n=78)	p-value
II	Physcial with Physical Functioning	0.65	<0.001	0.53	<0.001
III	Physcial with Role/Social Limitations - physical	0.43	<0.001	0.30	0.007
IV	Physcial with Bodily Pain	0.67	<0.001	0.46	<0.001
V	School & Play with Physical Functioning	0.67	<0.001	0.67	<0.001
VI	School & Play with Role/Social Limitations - physical	0.40	<0.001	0.50	<0.001
VII	School & Play with Bodily Pain	0.60	<0.001	0.48	<0.001
VIII	Emotional with Mental Health	0.48	<0.001	0.34	0.003
IX	Emotional with Self-Esteem	0.38	0.002	0.10	0.4

Interpretability

The OxAFQ was fast and easy to complete, with no obvious interpretation issues.

Detailed description is provided in the original *Paper I*.[105]

Outcome measures from study II, and III

Repeatability/precision for RSA measurements

Despite the challenges related to the marker distribution and the fragility of the cadaveric bones and stiffness of the tissue, overall good precision was obtained, and without systematic bias. The results were similar to a study on trapezoimetacarpal joint implants, where the precision was reported to be below 0.37 mm.[68] Precision of translation measurements was greater for the osteotomy (below 0.20 mm) than for the CCJ (below 0.24 mm). Rotations were not of clinical interest and are thus not reported in this thesis. The clinical precision of the RSA (*Study III*) resembled what was found in *Study II*, with better precision for translations than rotations (Table VI).

Table VI. Repeatability/precision results for RSA measurements of the osteotomy in *Study II* (cadaver) and *Study III* (clinical).

Study II. *Repeatability of n=20 double measurements of the osteotomy (dCF with reference to pCF). Combined precision for zero distraction, 1 cm, 1.25 cm and 1.5 cm distraction.*			
N=20	**Tx**	**Ty**	**Tz**
$Mean_{diff}$	-0.02	0.09	-0.02
SD_{diff}	0.06	0.10	0.09
95%CI	-0.05;0.01	0.04;0.15	-0.07;0.03
CR	0.11	0.20	0.17
Study III. *Repeatability of n=10 double measurements of the osteotomy; Combined for all patients.*			
N=10	**Tx**	**Ty**	**Tz**
$Mean_{diff}$	0.00	-0.03	0.03
SD_{diff}	0.03	0.05	0.07
95%CI	-0.02;0.03	-0.04;0.03	-0.02;0.08
CR	0.07	0.09	0.13

$Mean_{diff}$ is the mean difference between between double-examinations
SD_{diff} is the standard deviation of the $Mean_{diff}$
95%CI is the 95% Confidence interval of the $Mean_{diff}$
CR is the coefficient of repeatability, $\pm 1.96 * SD_{diff}$

Interim report of RSA measurements

The most interesting observation from the data comparison in *Study III* at each follow-up point was the proximal migrations of the dCF with reference to the pCF in the HATCP group, indicative of a compression of the osteotomy (Figure XII and Table VIII). The per-protocol interim results revealed a difference in compression at 6 months (HATCP Tx minus AUTO Tx) of 1.97 mm (99.8%CI: -1.65; 5.60). Sensitivity analysis yielded a one-sided 99.9% CI -2.08;6.03 and an intention-to-treat approach, including the reoperated patient yielded mean migration difference of 2.57 mm (99.8%CI: -1.65; 6.80). The migrations appear to cease at 6 months (Figure XII, *left*). Similar pattern is evident from the data on Ty (Figure XII, *middle*). Tz migrations were insignificant (Figure XII, *right*). As the 99.8% CI contains zero and the non-inferiority margin of 2 mm, it is not possible to make statistical inferences about the difference between the two materials. The total translations for each material are presented with the translations, indicating the total osteotomy migration.

Table VIII. Mean±SD(range) of Tx, Ty, Tz and TT in the osteotomy model within the two group at each follow-up point. Reoperated patient not included. 6 months follow-up was completed by 5 patients in each group. 12 months follow-up was completed by 5 patients in the HATCP group and by 3 patients in the AUTO group. *Adapted from Paper III.*

AUTO	**6 weeks**	**8 weeks**	**6 months**	**12 months**
Tx	0.14±0.63(-0.46;1.2)	0.18±0.72(-0.45;1.43)	0.25±0.96(-0.44;1.93)	0.21±0.99(-0.5;1.34)
Ty	1.44±0.99(-0.03;2.06)	1.33±1.06(0.13;2.27)	1.52±1.14(0.27;2.70)	0.48±1.63(-1.15;2.12)
Tz	-0.64±0.72(-1.6;0.14)	-0.60±0.76(-1.77;-0.008)	-0.61±0.49(-1.35;-0.18)	-0.17±0.27(.0.4;0.12)
TT[a]	1.47(0.14;2.87)	1.58(0.14;3.22)	1.81(0.46;3.58)	1.54(0.65;2.19)
HATCP	**6 weeks**	**8 weeks**	**6 months**	**12 months**
Tx	1.07±0.79(-0.24;1.76)	1.44±0.99(-0.28;1.13)	2.22±1.53(-0.21;3.61)	2.12±1.58(-0.33;3.65)
Ty	1.87±1.2(0.62;3.63)	2.37±1.40(0.85;4.14)	4.41±2.89(1.50;7.77)	4.12±2.58(1.55;7.29)
Tz	-0.70±0.96(-1.77;0.28)	-0.73±1.19(-2.40;0.60)	-0.49± 0.95(-1.96;0.99)	-0.45±1.2(-1.98;1.34)
TT	2.48(0.66;4.24)	3.14(0.90;4.81)	5.25(1.59;8.04)	4.96(1.68;8.20)

The excluded patient had migration data until 8 weeks follow-up: Tx=5.04 mm, Ty=3.09 mm, Tz=-1.13 mm (6 weeks) and Tx=5.82 mm, Ty=5.32 mm, Tz=7.83 mm (8 weeks).

[a]Total translation = $\sqrt{(Tx^2+Ty^2+Tz^2)}$

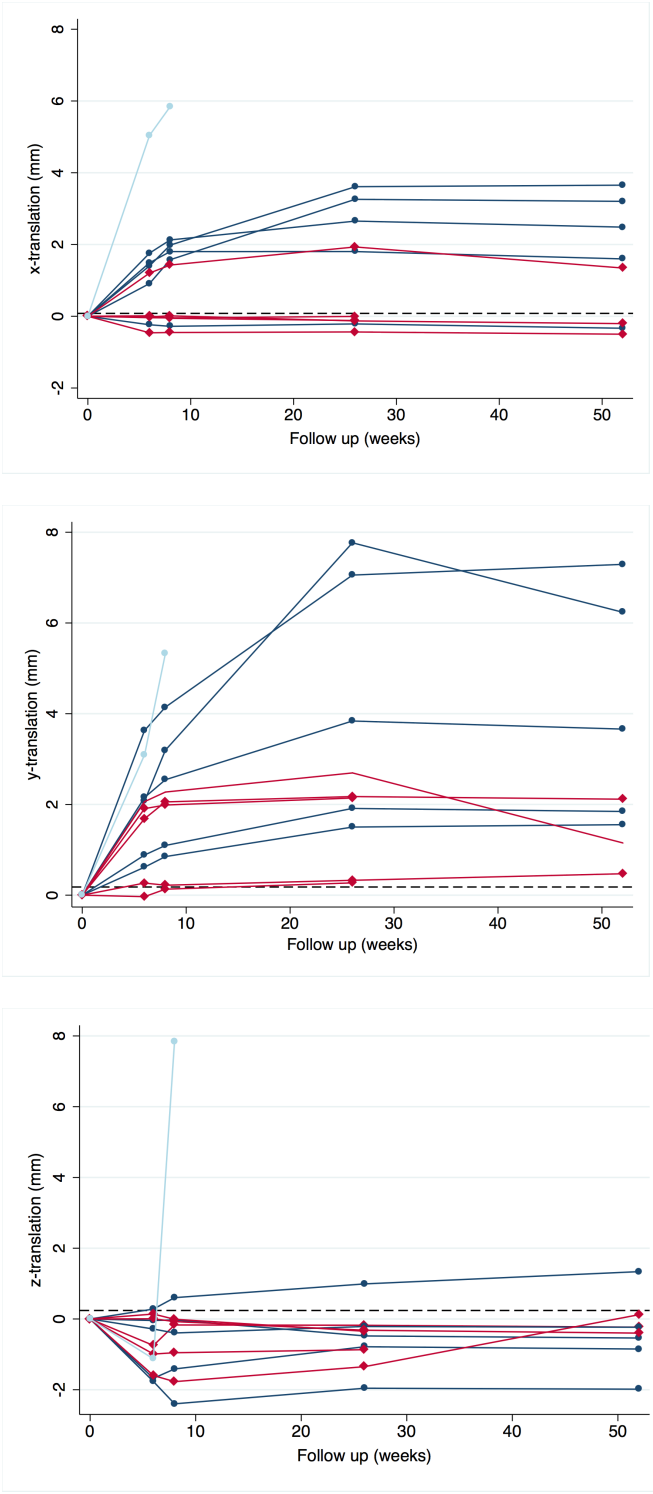


Figure XII. Translation of the dCF with reference to the pCF for each of the 11 patients. X-translation (left), y-translation (middle), z-translation (right). Primary endpoint is Tx at six months, but the 12 months follow-up are included for 8 patients in which follow-up was complete. The AUTO group is represented by red lines, and the HATCP group is represented by the blue lines. The light-blue plot represents the reoperated patient. *Adapted from Paper III.*

All surgeons reported signs of bone incorporation in the 6-week postoperative radiographs. It is noted that the radiographs were obtained with the patient casted.

Oxford Ankle Foot Questionnaire

With exception to the emotional domain, both groups reported improvement in functioning following surgery at six and 12 months follow-up (Table IX).

Table IX. Mean± SD domain percentage scores (including Footwear item) in the AUTO and the HATCP group. The reoperated patient is not included. *Adapted from Paper III.*

AUTO (n=5)						
Domains	N	Baseline	N	6 months	N	12 months
Physical						
Parents	5	45.8±22.4	5	79.2±23.8	3	87.5±21.7
Children	5	63.3±9.9	5	81.7±18.1	3	93.1±2.4
School & Play						
Parents	5	72.5±18	5	82.5±32.6	3	91.7±14.4
Children	5	82.5±14.9	5	88.8±16.8	3	95.8±7.2
Emotional						
Parents	5	80±20.4	5	83.8±23.6	3	72.9±46.9
Children	5	90±11.4	5	85±19	3	79.2±30.8
Footwear item						
Parents	5	40±28.5	5	70±44	3	66.7±57.7
Children	5	45±20.9	5	70±41	3	58.3±52
HATCP (n=5)						
Domains	N	Baseline	N	6 months	N	12 months
Physical						
Parents	5	32.5±12.6	5	55.8±20.3	5	70.8±17.9
Children	3	29.2±11	3	55.6 ±4.8	3	56.9±8.7
School & Play						
Parents	5	47.5±18.5	5	68.8±13.3	5	77.5±23.6
Children	3	43.8± 0	3	64.6±13.0	3	58.3±25.3
Emotional						
Parents	5	68.8±18.2	5	91.3±13.7	5	88.8±11.2
Children	3	70.8±7.2	3	91.7±9.5	3	93.8±0
Footwear item						
Parents	5	20±11.2	5	60 ±41.8	5	70±20.9
Children	3	25±25	3	41.7±14.4	3	50±25

Complications, pain and analgesics

All patients in *Study III* reported the postoperative donor site pain to be worse than the pain from primary surgical site. Table X summarizes the most important postoperative variables and the complications. Due to the small sample size data are presented as means and ranges. The mean pain at 24 hours was lower in the HATCP group but not significantly. Equal amounts of analgesics were administrated within the two groups. There were no differences in the total length of hospital stay or length of surgery.

Complications related to the primary operation field were encountered in the HATCP group at cast removal. There was an apparent inflammation at the surgical site in five out of six patients in the HATCP group compared to 1 out of 5 in the AUTO group (details are outlined in Table X). Donor site related hypertrophic scar (Figure XIII) occurred in three patients. Of these, two resolved at three months follow-up, and persisted in one patient. One patient, had signs of infection at cast removal (6 weeks), which did not respond to a 10-day oral penicillin cure. There was an apparent recurrence of the deformity. After 8 weeks follow-up it was revealed that the graft was colonised with Staphylococcus Aureus (SA) and Coagulase Negative Staphylococci (CoNS) (confirmed after tissue sample had been sent to culture and microscopy). The patient had the graft removed and received an intense and prolonged intravenous antibiotic cure with no succesive calcaneal lengthening during the duration of *Study III.*

Table X. Postoperative data for all patients included in interim-analysis (n=11). Data listed as mean(range). *Adapted from Paper III.*

Postoperative variables	**AUTO** (n=5)	**HATCP** (n=6)
Postoperative pain numeric scale (0-10)		
Baseline	0.8 (0-2)	0.7 (0-4)
24 hours pain, mean	2.8 (0-6)	1 (0-3)
24 hours pain, max	5.6 (3-8)	5.8 (0-9)
Morphine equivalents/kg body mass (mg, po)		
24 Hours	0.6 (0.3-1)	0.5 (0.2-1)
Discharge (total)	1.7(0.4-4)	1.8(0.4-4.4)
Total days of admission	2.8 (2-4)	2.6(1-4)
Duration of surgical time	1.7(1.4-2.2)	1.6(1.3-1.8)
Complications	**AUTO** (n=5)	**HATCP** (n=6)
Grade I		
Minor wound issues	n=3 (donor site)	n=2 (foot)
Grade II		
Antiobiotic treatment of wound infection (foot)	n=1	n=2
Grade IIIb		
Reoperation or wound revision		n=3

The gradings are based on the new and reliable complication classification system by Dindo et al [97].

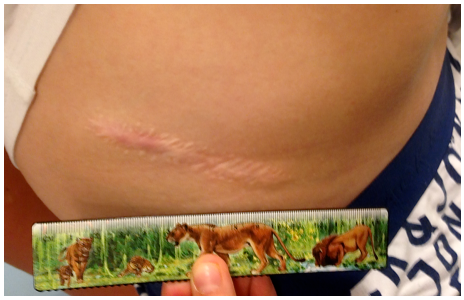


Figure XIII. Hypertrophic scar 2 years after iliac crest bone graft harvesting. Scar dimensions 8 x 1 cm. The patient complains of pain especially with jeans wear and is bothered by cosmetic issues.

Discussion

Oxford ankle foot questionnaire

The primary objective in *Study* I to assess the reliability and validity of the Danish version of the OxAFQ was achieved. The design of the validation proces emanated from the usual clinical setting in which the OxAFQ is intended for use. Tested in a typical population encountered in the Children's Orthopaedic Outpatient clinic, the OxAFQ was found to be free from systematic bias and with a random (intra- and inter) variation of less than 20%. The greatest random variation was found in the footwear item (Table IV).

The physical domain scores were consistently lower than the scores of school & play and emotional domains. Overall the mean scores and the random variation in this study population concur with the developers's findings.[82]

Internal concistency expressed as Crohnbach's alpha was overall high, albeit lower in the emotional domain. Noticably item 14 ("Has anyone been unkind to you because of your foot and ankle?") was usually scored higher than the other items in the emotional domain as it reflects a "bullying issue". This should be kept in mind when interpreting the scores in future studies. The question has great value, as it can reveal an important issue that might be overlooked, if the child fails to bring it up itself.

The participants found the OxAFQ fast and easy to comple, providing minimum patient and consultant burden. Ceiling effects were observed in the school & play and emotional domains.[105] These reflect the majority of patients who in general reported high scores in the physical domain, as they were discharged from further follow-up by from the surgeon after the first questionnaire administration.

Paradoxially, foot and ankle issues are reported to be the most predominating among musculoskeletal disorders,[106] yet in the population encountered at the Outpatient Clinics at Children's Orthopaedics, it did seem to affect the patient's ability to participate in physical activities (Table II). Possibly, the children are not as physcially limited by their condition, that it hinders them to participate in activities. Rather, they adjust to their situation and do what they are able to. Perhaps, parents, as well as the general practitioner are insecure about how to manage the foot/ankle disorder and therefore seek the advise from a specialist. The emotional domain was mainly affected in patients awaiting elective surgery as we would

expect, since surgery is reserved for the more severe cases. In line with previous findings[82,107] the children reported less impaired quality of life based on the OxAFQ item scores than their parents, which was statistically significant in the emotional domain (p=0.02).

It is important to recognise the value in obtaining both child and parent(s)'s perspectives on health related quality of life, as their goals and expectations to the treatment may differ. To understand the possible variables, which account for well known discrepancies between child and parent (proxy) reports of health related quality of life, readers are encouraged to read the interesting review on this topic by Eiser et al.[108]

Having found OxAFQ valid an feasible to assess health-related quality of live in children foot and ankle, it was used in *Study III* to compare the outcomes of the calcaneal lengthening osteotomy in children who recieved HATCP and AUTO graft.

The baseline OxAFQ score distribution in both groups resembles those reported by previous authors who have investigated HRQoL by using OxAFQ in paediatric pes planovalgus.[107,109] Kothari et al.[107] did an excellent study on the health-related quality of life in children with typically developing feet (TDF) and those with paediatric flexible flat feet (PFF) benchmarked by the PedsQL™4.0 domain scores. All domain scores, particularly the physical domain and the footwear item, were lower in our study population compared to the TDF and the PFF. This is not suprising, since we would expect elective patients to be severely affected by their foot pathology. Interestingly in a study by Martinelli et al.,[109] the Italian OxAFQ was validated in 61 children who were scheduled for surgical correction of flexible pes planovalgus. The preoperative scores were similiar to those of the PFF population in the study by Kothari et al., and thus less severely impaired than those of our population. This raises an interesting question on how surgical indications are determined in different cultures.

It may seem that the HATCP group had lower baseline scores compared to the AUTO group, however both groups reported similar improvement in the OxAFQ domains scores representing physical function, participation in everyday life activities, emotional well-being and footwear use.

Radiostereometric analysis

Inspired by the field of hip joint replacement surgery, where RSA measurements have become a gold-standard to assess the efficacy of new implants, owing to the high accuracy and precision of the system, it would be favourable to use RSA as measurement tool to assess stability of the CLO in paediatric patients (*Study III.)*

Study II was designed to investigate the feasibility and precision of RSA on CLO, as there were no papers in the literature for this application of RSA. The guidelines as proposed by

Valstar et al.[95] were followed and the RSA set-up and data analysis were standardised to ensure, that the RSA set-up would be valid for the envisioned purpose. Besides the requirements for the radiographic equipment including calibration cages and software, the surgical feasibility needs to be tested by those surgeons who plan to use the RSA.

Methods of marker-insertion, -number, -stability and -distribution are important to manage in order for the accuracy / validity of the subsequent migration calculations.
During *Study II* we learned that the osteotomy site could be used to access the distal (anterior) and proximal (posterior) calcaneal bones, which made marker distribution feasible. Insertion of the markers in the cuboid required expanding the intial skin incision distally, which would also create a slightly longer scar than usual. Drilling small holes prior to bead injection protects the bead injector from wear. I would recommend using different sized markers for neighbouring small bones, as this makes it easier to distinguish between the bone models during marker-based software analysis.

Above all surgeons who are new to RSA and plan to use it in future studies are encouraged to conduct cadaver experiments, as this more accurately reflects how the instruments works in a clinical setting. In this set-up, the foot was positioned with the lateral side facing the calibration cage, the medial side facing upwards and the ankle in neutral. Using this supine foot position in future studies allows for easy comparisons of migrations.

Having established the feasibility and precision of RSA for measurements of calcaneal lengthening osteotomies in cadaver feet, the next step was to use it as a primary outcome measurement in *Study III.*

Structural durability

Structural durability of bone grafts is reflected by a stable osteotomy in an asymptomatic patient. If these conditions were present to a similar extent in the HATCP group as in the AUTO group, the HATCP woul be a feasible alternative to autologous iliac crest bone graft harvesting. Fundamentally the new bone graft material must meet the main requirement, which is to function as a stable scaffold allowing for new bone formation to occur within a reasonable timeframe. In the end, stability is a virtue.

The AUTO group reached stability of the osteotomy at six weeks after surgery while the HATCP group had achieved stability at six months. The HATCP group lost a mean of 1.97 mm (99.8% CI: -1.65; 5.60) lengthening more than the AUTO group. By including the patient with infection where follow-up was terminated after eight weeks, the migration difference was at 6 months was 2 mm (99.8% CI: -1.65; 6.80). The predefined noninferiority margin was exceeded and the CI limits were skewed towards higher migration values indicative of osteotomy compression. The dorsal migration of the distal calcaneal fragment

in the HATCP group, which occured in parallel with the distal migration is supportive of delayed bone incorporation. These findings of delayed incorporation of HATCP are in line with two other recent clinical studies conducted on calciumceramics. In a feasibility study by Balakumar et al[44] triphasic calciumceramics (HASi) were used for different purposes, including CLOs (no information on fixation provided) on 40 patients. Based on radiographs from 35 patients, they reported complete healing in 19 cases and partial healing in 16 cases at six months follow-up. The mean incorporation time was overall longer for the autograft than the HASi graft, with six months for cancellous defects (such as calcaneus), and longer for cortical defects. They reported one case of malunion and four cases of non-union. Their primary outcome measure was graft incorporation at different follow-up time points (3, 6, 12, 18, 24 months after surgery) assessed by two different observers by conventional X-rays. They reported an inter-observer reliability by a weighted Coehen kappa of 0.82 (95%CI: 0.55-1.1). Graft incorporation was defined as "trabecular continuity throughout the graft, with loss of radio-opacity and gradual blurring of the sharp margins of the HASi block. If such changes were noted only on the surface of the block, it was classified as partial incorporation." To rely on a qualitative/subjective measurement method for evaluation of a new treatment limits the validity of the study results, and also there is a risk, that the observer could be biased in either direction. The HASi bone graft has a visible appearance wherefore blinding during the union-assessment is not possible. Although blinding was not possible in *Study III* either, bias cannot be imposed on the software analysis as the migrations numbers are generated from matching the defined bone marker models.

Glazebrook et al[110] conducted a pilotstudy on the performance of B2A-coated ceramic granules in various foot and ankle surgical procedures. Their primary outcome measure was bony fusion at six months assessed by CT-scans. A second CT scan was obtained at nine months and conventional X-rays were obtained at 6 weeks and 3, 6, 9 and 12 months. They defined radiographic fusion as greater than 50 % joint fusion. CT scans are superior to conventional X-rays in assessment of graft incorporation.[111] The main limitation is their beam-hardening artefact, and the high radiation dose. For osteotomy stability measurements we favourise RSA instead of CT, as we can perform multiple RSA examinations to assess bone incorporation over time, without exposing the patient to unnecessary high radiation doses.

Pain

Patients who received AUTO graft graft reported the hip pain as worse than the foot pain. Yet, with this small sample size there was no significant difference in pain scores between the AUTO and the HATCP graft. The method by which pain was measured is liable to various sources of bias: The postoperative medication regime was adjusted according to the individual patient, due to unexpected events such as non-functioning tibialis cathether,

suboptimal casting that exhibited pressure on bony prominences, pain related to intravenous procedures and different sensitivity to analgesics. Also, from clinical experience and previous neuroscientific research[112], it is known that anxiety (e.g. being in an unfamiliar enviroment), anticipation of painful procedures can increase the indiviual perception of pain. These findings are in accordance with those from a recent prospective study on donor site morbidy following anterior iliac crest bone grafting in 33 children (34 hips) and adolescents [113]. VAS score was used to differentiate between the pain at the donor site and the recipient site. A strong positive correlation between the measurements at the sites was established. In addition mean VAS was significantly higher in patients who underwent a flat foot procedure than other corrective osteotomies. The impact of the primary procedure on acute pain seems to be greater than the bone graft harvesting itself. Clarke et al.[113] found no complications to graft harvesting, except for a lateral cutaneous femoral nerve injury (numbness over lateral thigh) in a single hip, which resolved at three months. In this thesis only one patient had persistent hip pain at final follow-up, due to hypertrophic scar tissue. In agreement with Clarke et al.[113] it appears the morbidity associated with iliac crest bone graft harvesting might be generally overexposed. Instead the perspective should be aimed towards improving other strategies, such as optimising the pain management, especially in patients who are sensitive to postoperative pain. This topic will be discussed further in the Perspective and Future Research section.

Complications

Few clinical research papers are available on the safety and feasibility of structural calcium ceramics in foot surgery.[44,110] Glazebrook et al.[110] reported 3 out of 12 wound issues (versus none in the autograft arm (n=12)). However, they did not attribute these to the graft material. We observed five out of six wound irritations, of which one case was a Staph. Aureus and CoNS infection in the HATCP graft versus one out of five wound irriations in the AUTO group. Staph. Aureus is a very common and feared organism that accounts for infections related to orthopaedic surgery.[114] The ability to create biofilm makes it resistant to systemic antibiotics. CoNS are oppotunists, also associated with implanted foreign bodies,[115] as implants are suspectible to bacterial infection, because of the "locally compromised host defence".[114] The brittle nature of the HATCP graft made it challenging to prepare and insert in the osteotomy, as the graft easily dissipates into the surrounding tissue and is difficult to wash-out. Vertical cracks can occur during the insertion, thus necessitating repetitive and careful lengthening attempts. The grade II and IIIb complications observed in the HATCP group might be related to the dissipation of the material into the surrounding tissue. However, the question of whether the severety of the Staph. Aureus and CoNS infection would have been different in an autologous bone graft remains unanswered. Either way deep infection would require graft removal.

Methodological considerations

Oxford ankle foot questionnaire

The greatest challenge in designing *Study I* was how to assess construct validity. Morris et al.[61] compared trauma patients with elective patients, and also compared the OxAFQ scores with those of the Kidscreen questionnaire. It had been considered to administrate the OxAFQ to trauma patients, but was refrained from for two major reasons. Ultimately the questionnaire should be tested in the population in which it is intented to be used. Trauma patients have issues of acute and straight-forward character and are all dealt with immediately in the emergency room (ER). As the ER and the Dept. of Children's Orthopaedics at AUH are seperate departments it was decided not to include trauma patients.

In *Study I* the construct validity was assessed by nine a priori hypotheses. Hypothesis No. I was set up to investigate if the questionnaire was capable of discriminating between disease severeties (discriminant validity). Therefore patients were asserted into three groups depending on their follow-up status in the outpatient clinic; the newly referred (group 0), those attending clinical follow-up (group I) and those referred to surgical intervention (group 2). It was hypothised, that the elective patients would be more severely impaired in their physical performance than those attending the clinic for follow-up of an ongoing treatment, which was confirmed (children p=0.0002, parent-proxy p=0.03). Another more accurate approach would have been to allocate patients based upon the severety of their diagnosis. However, the graduation of the diagnosis is subjective, while dividing the patients in groups based on their visitation status is more reliable, but less accurate. A possible method could have been using a visual analogue scale (VAS) approach, asking a question like "To which degree overall do you find your foot/ankle issues is a problem to you?" Capturing the answers of the parent and the surgeon would provide more information to subsequently assign the results from the VAS score to the proposed gradings of mild, moderate, severe.

The newly referred patients to have the highest scores, overall. Possible explanations for this may be that the general practitioner often refers patients with issues of a

benign character, due to lack of knowledge on the subject matter, or the foot/ankle issue is a recent observation, which has not yet deteriorated.

It is always necessary to consider the validity of the stated hypothesis. Eight hypotheses were designed on positive correlations between domains which were expected to measure similar constructs (convergent validity) between the OxAFQ and the CHQ. These hypotheses were confirmed with regard to the child version and six out of eight in the proxy version (Table V). There are several implications to using the CHQ as a comparator instrument. When comparing a disease-specific PROM with a generic PROM, it is not fair to assume more than a moderate correlation between similar domains. This issue has also been faced by Kocher et al[86] in the validation proces of a Danish knee-specific HR-PROM. We chose the CHQ as it was the only thoroughly validated HR-PROM at the time of study design. Optimally, the administered versions of the CHQ should have been of equal length versions to avoid potential bias. However, there was no evidence of clinically relevant discrepancies between the child and proxy scores in the hypothesis testing.

The construct validity with regard to the hypothesis testing using CHQ as a comparator instrument, was not assessed in children younger than nine years (13 children). The Kidscreen questionnaire could have been used for subgroup analysis, altough the Danish availble translation had not been validated at the time of study design. On the other hand, Morris et al.[81] suggested, that answers from children below eight years should be interpreted with caution. Evidence from the psychiatry litterature, suggest that the younger and the less cognitively developed children, the less reliable are their responses to self-administered questionnaires.[116] In conjunction with clinical experience, parent (proxy) answers would be considered more reliable than those of the younger children.

The retest questionnaire should optimally have been completed in the same setting as the done initially. Eventhough the participants were instructed to complete the questionnaires seperately, the retest data demonstrate more child-parent coherency in the responses. If one should be very meticulous about the systematic bias, one could think that the parents adjusted their responses to school and play and emotional domains after communicating about this with their children at the retest. On one hand, the small sample size in this study warrants caution for relying on this interpretation. On the other hand, if true, important observations can be drawn from this methodological shortcoming - the OxAFQ serves as a facilitator of the communication between children and parents on important issues, which the child and parent have not yet considered or talked about.

A conservative ICC (1.1) was assessed, which may explain the lower ICC found here than previously by the developers, who used an ICC 2.1 approach. Without going into an in-depth statistical discussion of the ICCs, there are a few important things to consider when using ICCs to investigate the results obtained with a measurement instrument. The ICC is a relative measure of reliability, used to distinguish subjects/groups from each other.

ICC by itself cannot be used to disclose a systematic error[2] in the measurement instrument. Thus, from the score results of the OxAFQ the variance in a homogenous group would be expected to differ from the variance in a heterogenous group, as it is more difficult to distinguish subjects from each other if they are relatively similar, than if they are very different. The most sound approach would have been to calculate ICC within groups expected to be homogenous instead of the entire study population (which is heterogenous in our study). However, as the primary statistical approach to measurement error followed the Bland-Altman method, this was not pursued further, but I do take it into consideration for future research.

Adding the ICC to the Bland-Altman approach of measuring the relative and absolute measurement error with regards to reliability, may seem redundant, however it allows the readers to more easily compare data with those of the developers.

Other measurement properties such as the content and face validity, are usually validated based on subjective judgments by an expert panel and in current study these properties were judged to be appropriate by the senior surgeons of the department.

Testing responsiveness was not within the scope of this thesis, but we are running an ongoing longitudinal validation of the OxAFQ. Strictly, if a PROM is to be used as an outcome measure to assess changes due to interventions, the responsiveness must be assessed. However, one could argue that if the questionnaire is able to differentiate between disease severeties, then it would also be assumed to detect a change over time, as longitudinal validity is principally the same as discriminant validity in this case, but measured over time in the same subject.

[2] However, a standard error of mean (SEM), can be derived from the ICC by the formula SEM=SD(observed scores)*√(1-R).

Radiostereometric analysis

Study II was designed to investigate the precision and feasibility of RSA measurements of calcaneal lengthening osteotomies in preparation for *Study III*. Paediatric cadaver feet would have been the ideal phantom, but for obvious reasons these were not possible to procure. Stiff, adult cadaver feet have a bigger anatomy and greater joint rigidity. A greater effort is required to perform distraction of the osteotomy without fracturing the fragile bones. Fresh frozen cadaver feet may have been easier to work with, but again difficult to procure. Despite the stiffness of the tissue, the distraction and bead insertion were still feasible. Accuracy was not within the scope of *Study II*, due to the obvious differences between adult cadaver (non planovalgus) feet and vital paediatric (planovalgus) feet. This would also require baseline measurements in zero distraction of the osteotomy in the clinial setting prior graft insertion, which was not possible with the current marker based technique. The precision of the RSA measurements are based on double-examinations in range of distractions from 1 cm to approximately 1.5 cm, which is the usual graft lengthening done in the clinical setting. A different approach would have been to take e.g. 10 double-examinations in the same position, without compromising the validity of the results. However, using different graft settings reflects the clinical scenario, with small size variations in the bone graft used. Also using several cadavers for bead insertion and RSA examinations enables to assess the feasibility of RSA for calcaneal lengthening osteotomies.

The external validity of *Study II* is compromised by the characteristic differences between cadaver feet in the present study and those of paediatric patients with pes planovalgus. However, as the operative procedure and beadmarking were carried out similarly to how it is actually done in the field setting, and the stiffness of the cadaver feet challenged the procedure, RSA would be expected to be more feasible in the target population.

Moving on from the RSA set-up *in vitro* to *in vivo*, *Study III*, markers were inserted into each side of the calcaneal osteoteomy to allow for measurements of the osteotomy stability, which would represent the stability of the graft material. As bone growth is still potentially ongoing in this study population, markers were not inserted in the apophysis of the calcaneus, axial bone growth would not be expected to interfere with the marker positions.

The most important issue in using bone models for osteotomy stability is tha we measure the center of the marker configuration (rigid body), which is determined by the geometrical shape of the created bone models. We are confined to measure relative motion between the bones in the context of the landmark positions. Therefore accurate measurement of graft compression/length cannot be obtained, but it provides valuable information on the relative magnitude and direction of the motion of the osteotomy. Patient positioning was standardised by using the cast as template during each follow-up. Although it may seem indelicate, it was feasible and the patients made no objections to the set-up.

From the migration results in Table VIII and Figure XII, it appears the bone graft has not obtained complete stability at 6 weeks in the HATCP group. However, the X-rays obtained at 6 weeks which were the basis for cast removal had been described to show signs of union in the graft-host interface. By routine practice at our department six week postoperative X-rays are obtained while the foot is still casted. To assess bone bridging per se is exposed to subjectivity.[97] Adding a cast does not make it easier. Therefor this approach to should be questioned, at least when using HATCP bone graft material.

Advocates of radiological angles would disapprove of the lack of weight-bearing X-rays as part of the follow-up protocol to measure deformity correction. Yet as the primary outcome was lengthening maintenance measured by RSA, it was not considered necessary to add on extra X-rays. As mentioned previously, conventional X-rays were only indicated preoperatively to confirm the deformity. Surgery is not planned based on the X-ray, but rather on an individual assessment of each patient based on clinical examination and symptomatology. However, to allow for some external comparison of the study population three radiographic angles were assessed to give an overview of the radiological perspective of PPV deformity: the calcaneal pitch angle (CP) and the talo-first-metatarsal angle (Talo1M), which are both suggestive of the collapse of the longitudinal arch and the TC angle which reflects the hindfoot eversion and valgus position when increased. The CP did not differ from normal values, and the Talo1M was increased (Table III).

Pain

To measure pain in paediatric patients is a difficult task. The interest was to compare overall pain between the two groups. The AUTO group was expected to experience a greater max- and mean pain during the first 24 hours after surgery. However, in such a small sample size it is difficult to control for some extent of information bias. E.g. one child in the HATCP group experienced pain and anxiety related to intravenous procedures and thus reported high pain. Three patients in the HATCP group, had issues with the cast being tight and thus exerting pressure on bony prominences, that was recognised on the second postoperative day. It is important to realise, that when pain management is adjusted to each patient's need, then we actually measure the effectiveness of the treatment, not "true" pain.

Furthermore, children are prone to behavioural bias, depending on the emotional condition they are in at the time of the pain scoring. All participants in the AUTO group agreed upon the hip pain to be the worst. However, based on opend-ended questionning with the children and parents at the follow-ups, it did not interfere significantly with their daily function after dismission from hospital. Only one patient still reports chronic pain related to hypertrophic scar tissue at the hip (Figure XIII). Optimal managament of hypertrophic scar tissue is a research area which is still under investigation[117] and perhaps when an adequate treatment will be provided, then donor site morbidity will be less of an issue.

The method of measuring pain in this thesis resembles that of Clarke et al.[113] The authors differentiated between hip pain and pain from the surgical site, by asking patients to rate each site by using a visual analogue scale. In our study, patients were asked whether they were in pain, and if yes, then localise the pain and rate it. In case of autograft harvesting, they were asked which pain was worst. This approach was chosen, because it was simple and in line with what the children were expected to cope with, as the pain assessment was done at frequent intervals during admission. It was not of interest to grade the hip pain versus foot pain, but there was no overall difference between the mean pain in the two groups. It is known that surgery via pain signals induces a general inflammatory response, which would also be expected to add on the postoperative pain.[118] In line with this phenomenon, Clarke et al.[113] found a strong correlation between the pain at the recipient site and the hip. A limitation in our method and neither reported by Clarke et al. is, that only pain at rest was reported, not supported by asking to the dynamic pain. This may underestimate the pain score in the AUTO group, as the patients do exhibit greater pain in relation to mobilisation (such as getting out of bead for bathroom visits) when they have a wound at the hip. It is possible that an objective pain scale[119] would have strengthened the pain assessment/measurement, but due to lack of resources it was not integrated in the study protocol.

Noninferiority design

The design of *Study III* warrants a seperate methodological discussion. This study was initiated by one principal question: Can we use HATCP graft with similar stability results as AUTO graft in calcaneal lengthening osteotomies in paediatric patients with symptomatic flexible flatfeet and thereby circumvent the donor site related morbidity? It is obvious that this formulation is composed of several questions. What is similar stability results? What is the most appropriate way to measure it? And how do we measure donor site morbidity?
The investigation was conducted as a randomised noninferiority trial. By using a "through the looking glass" approach,[120] the working hypothesis was: HATCP is noninferior to AUTO graft by no more than 2 mm compression of the osteotomy. The obvious strucural differences between the autograft and the HATCP graft makes it highly unlikely that HATCP would be shown superior to AUTO graft, and thus a null-hypothesis of no difference in outcome between the two methods would not be appropriate. Supported by a previous animal study[51] on similiar composition of HATCP graft and the brittle nature of the material, we expected the bone incorporation into the HATCP graft would require longer time and a certain amount of osteotomy compression would have to be accepted. The expected benefits of HATCP graft would be to completely avoid donor site morbidity and most importantly to reduce the postoperative hip pain. As the aim was to show that the HATCP graft was not

unacceptably worse than the AUTO graft, the most sound approach was to conduct a noninferiority trial.

The next challenge was to define a reasonable noninferiority margin. The appropriate correction is performed at the surgeon's discretion with typical measurements of the lateral border of the graft ranging from 1-1.5 cm. Based on the senior surgeons's clinical experience, up to 2 mm loss of lengthening was considered acceptable. Critical statisticians might not approve the justification for choosing the margin based on sound clinical judgement instead of statistical approaches, and although we are dealing with bones and not drugs, we do recognise this limitation. The choice of noninferiority margin is essential for determination of the sample size, and should preferably be based on extrapolation of treatment effects of historical superiority trials.[120] However, as no previous study has quantified osteotomy stability, using sound clinical judgement in this specific case is a justified as statistical reasoning.

Noninferiority designs are suspectible to lack of "assay sensitivity",[120] an issue which was present in this trial and which have complicated the interpretation of the interim results. Assay sensitivity reflects the property of a trial to "distinguish an effective treatment from a less effective treatment".[120] Without this, a new treatment can be shown noninferior if the standard treatment is not shown superior to placebo. In *Study III*, the AUTO graft, which was expected to be stable, did undergo a significant compression in one case, and therefore the HATCP stability results demonstrate a likelyhood of noninferiority.

Another important lesson was obtained from inviting patients to participate in the RCT. Nine of twenty eligible partiticpants declined to participate, with the main reason being that parents found it very hard to cope with the randomisation proces. A parent noted: "One thing is providing my own body for the greater purpose of Science - I would gladly do so - but another thing is giving up my parental responsibility and leaving the treatment of my child up to chance." This issue has been previously described[121] and faced in other randomised trials where paediatric patients comprised the study population.[121,122] As the outcome measure is osteotomy stability measured by stereoradiographs, a prospective cohort study would have worked as well.

Interim guidelines and stopping guidelines

To date, there are no unequivocal guidelines for how to properly perform an interim analysis, but I will try to shed light on the important issues, that are considered essential for the validity of the analysis. In this specific arcaine field of statistics, there is an ongoing discussion on how to choose the appropriate method - which requires statistical in-depth knowledge. Therefore the methodological concerns will be approached from a clinical point of view, rather than a statistical, as this is beyond the scope of this dissertation.

In clinical trials where new materials are to be implemented, careful data monitoring is required in order to detect and minimise patient exposure to unexpected or toxic outcomes.[123] In addition unnecessary continuation of studies to prove benefits of the given intervention is ethically irresponsible. With respect to ethical concerns that follow any implementation of new bone graft materials in children, one predefined interim analysis was planned to be conducted halfway through *Study III*. Optimally, the data analysis should have been done by a data monitoring committee (DMC). This should be an independent and multidisciplinary commitee (including epidimiological, clinical, statistical and ethical expertice), which task is to provide unbiased monitoring of the accumulating evidence (clinical, as well as statistical). Ultimatly, the role of DMC is to advise the trial organisers/investigators who bear the ultimate responsibility for the conduct of the trial.[123] In order for a DMC to have credibility, the members should not have any relations to the trial, the should be multidisciplinary, possess qualifications to operate in a DMC and know-how within the specific research area. Unfortunately, there were no available resources at the time of trial inititation to assemble such a committee. Lack of a DMC is a major limitation to this study protocol. The entire data analysis was performed by the primary investigator, and the synthesis of data, stastitical considerations in conjunction with the feasibility of the material in the operating theatre provided the foundation for the decision made by the senior surgeons's of the department regarding the future of the study.

The interim analysis is based on considering a two-sided 99.8% CI of the Tx together with the other predefined stopping guidelines. If the upper limit does not exceed the noninferiority margin, then noninferiority of the new treatment has been proven. The methodology used in this interim analysis was proposed by the statisticians, consulted for the design of the study. The idea behind this approach, is that small studies should be backed by significant statistics in order to remain credible. Other, less conservative approaches could have been applied, but it was decided to rely on those proposed by our statisticians.

Other stopping guidelines, such as the futility index[124] could have been to proposed. Without going further into the statistical process, the futility index can be utilised to assess the probability of a trial on an innovative therapy to yield succes if the small accumulated data suggest that the new treatment is inferior to the standard treatment[124]. As a principle, it is crucial for the study validity, to predefine stopping guidelines and for which variables they apply. It is important to acknowledge that the primary function of statistical reasoning is to serve as guidelines, not as termination rules.

Some of disadvantages in stopping a trial early, within the context of this study, include such as lack of credibility due to small sample size, imprecision (wide confidence interval for the "treatment effect"). The paper Education and Debate: When to stop a clinical trial?" by Pollock provides an excellent overview of the generalissues in stopping a trial early. [125]

Standing at the crossroads

All the forementioned considerations leads us to the greatest ethical dilemma in this trial: To stop or not to stop further patient inclusion? On one hand we face the statistical descriptions of the primary outcome - the stability of the osteotomy. The robustness was checked by performing both per-protocol and intention-to-treat analyses. In either case the disturbing finding of a mean difference and high upper limit (up to 7 mm) exceeding the predefined noninferiority limit was suggestive of the doutbful osteotomy stability in the HATCP group. We recognise that although the data might be pointing towards inferiority, they statistically inconclusive. On the other hand we consider all the expected and unexpected additional outcomes and find ourselves to suspect the new treatment to be worse. We discussed whether to stop inclusion before the evidence reached "statistical significance in the wrong direction"? However, in light of the less encouraging migrations numbers, feasibility challenges and wound issues in conjunction with the fact, that the donor site pain is not as incapacitating as anticipated, we do not believe we can justify to continue the trial. Ethically, if the *new treatment is suspected to be worse, then weaker evidence is required in favor of stopping the trial*.[125]

Generalisability issues

The overall generalisability of the results of *Study III* is limited to a minor extent by the lack of an unequivocal set of guidelines for how to classify and manage PPV.

The study population comprised idiopathic, JIA and mild neuromuscular PPV (cerebral palsy), therefore we cannot know the nature of bone incorporation in HATCP graft in a more severely affected patient. The characteristics of our study population was described using a range of measurements methods. The validated OxAFQ was used to address HRQoL, as a painful and disabling PPV is the most important indication for surgery. The arch index was assessed to provide an anthropometric description of the longitudinal arch. Overall the mean value ±SD was 0.3 ±0.1, suggestive of a flattened arch in comparison to a healthy subject, which is based on data derived from a cross-sectional study on 7788 healthy children aged 1-13 years, with an arch index of 0.2±0.1.[126] As mentioned the baseline radiographic measurements were suggestive of an overall relatively mild PPV deformity.

Conclusion

This thesis was initated by the following question: *Can HATCP graft be used with similar stability results as AUTO graft in calcaneal lengthening osteotomies in paediatric patients with flexible flat feet and thereby avoiding donor site related morbidity?* Similar stability results would be assessed by measuring the health related quality of life in relation to the foot and by assessing the stability of the osteotomy using valid tools. The validation of these tools was the first two steps.

The Oxford Ankle Foot Questionnaire was translated to Danish, and it was found feasible and valid for use in 5-16 years olds, it is advisable to use the proxy version for children younger than nine years.

Overall RSA was found to be a precise and feasible measurement instrument which allows for three-dimensional migration measurement of the stability of the osteotomy in children who undergo CLO for pes planovalgus.

Stability of the osteotomy was found to occur at a faster rate in the AUTO group (6 weeks) than in the HATCP group (6 months), thus far more lengthening would be sacrified in the HATCP group than we would accept. From a clinical perspective, the changes in the OxAFQ scores were similar in both groups. Safety was compromised in the HATCP group by feasibility issues related to brittle nature of the material and to wound inflammation.

In conclusion HATCP graft is of limited value as a structral bone graft in CLOs when no fixation is applied and AUTO graft remains gold standard in calcaneal lengthening osteotomies at our facility.

Perspectives and future research

Ultimately, the overall perspective of this thesis is to improve the surgical outcome in patients with symptomatic flexible pes planovalgus. To do that, it is crucial to understand how it affects them it in their daily lives and to assertain that we use valid measurement instruments to assess the surgical, as well as, clinical outcomes.

Oxford ankle foot questionnaire

In conclusion, I encourage every paediatric orthopaedic surgeon to incorporate the OxAFQ within their tool kit for three major reasons. First, it is valid and reliable, it can distinguish between disease severities, it measures what it is intended to measure, it samples relevant domains in children with foot and ankle disorders. Second, it is easy to understand and expedient to complete. Third, it has the potential to assist the surgeon to elucidate important aspects of foot/ankle symptoms, and manage the expectations of both child and parent.

The discriminative validity of the OxAFQ was assessed and not the responsiveness. This is also important to establish for future evaluative purposes. As the OxAFQ provides a snapshot of the impact of the foot/ankle problems (during the last week) on quality of life, any recent lower extremety trauma should be noticed, as this is a potential limitation to the validity of the response if used for evaluative purposes.

It would be interesting to establish normative values for OxAFQ in children with pes planovalgus in order create cut-off limits of disease severety. Kothari et al.[127] took the first step, but larger studies are needed including patients from the schools, primary and secondary healthcare.

Interpreting the OxAFQ scores by items instead of domains could yield more valuable and accurate information, as the flatfooted child belongs to a heterogeneic population. We know the foot dynamics differ between children with normal developing feet and children with flat feet and we know that the quality of life is impaired in children with flat feet.[107,128] Kothari et al.[127] recently showed, that greater impairment of the physical domain of the OxAFQ is associated with a greater degree of maximum hindfoot eversion

and forefoot supination using three-dimensional gait analysis and the Oxfort Foot Model. Three dimensional gait analysis is a very time consuming set-up and the marker placement on the skin is crucial for the accuracy of the data output. In complex cases a high-technology gait analysis may provide additional information, which could not otherwise be accurately obtained in easier way. In the particular case of flatfeet with absence of severe underlying pathology, one should consider whether high-technology gait analysis would add any particular value. Most likely a thorough clinical gait analysis would be the most appropriate approach.

Radiostereometric analysis

The development within optimisation and feasibility of RSA is an ongoing focus of research, and today fluoroscopy-based markerless RSA systems can enable *in vivo* analysis of joint kinematics during gait.[129] There is an abundant room for expanding our understanding of the foot joint kinematics before and after any surgical procedure, and hopefully optimising current treatment strategies. In the future, coupling plantar pressure sensors with dynamic RSA examinations before and after interventions would provide valuable information on the kinetic and kinematic changes in relation to surgery. Specifically for the CLO, there is a concern for the subluxtion of the CCJ upon distraction and therefore it has been advocated to stabilise the CCJ.[11] However, no study of high methodological quality and with long-term follow-up has shown the benefits of stabilising the CCJ. This would be an interesting and relevant question to uncover.

Cost-effectiveness

As with all new techniques/materials prior to their regular implementation in clinical practice, a cost-effectiveness analysis should be undertaken to support their feasibility.

The bone incorporation occured at a slower rate with the HATCP graft compared to tricortical autograft, but the osteotomy was stable at 12 months in both groups. As a substantial amount of lengthening was sacrified, HATCP has limitations when used by itself as a structural bone graft. However, fixation of the osteotomy without pentration of the HATCP graft could perhaps prove better results. Extending the immobilisation period by

two weeks of partial weight bearing may win some time for union between graft and host-bone.

Future inquisitive research should focus on the microstructure of the new bone that is produced in response to creep substitution when using HATCP graft versus tricortical iliac crest autograft after longer follow-up. This information can be obtained by microCT technology (HR-pQCT, high resolution peripheral quantitative computed tomography).[130]

Graft failure remains a liability. Therefore it is important to elucidate on potential risk factors for delayed union or non-union for each patient in the operative planning. Epidemiological studies on all patients who have received bone graft could contribute to fill out this knowledge gap.

Instead of searching for other bone graft materials we need to optimise the donor site management. Careful surgical technique is alfa and omega to obtain a good result. More obscure issues such as donor site pain and developement of hypertrophic scar tissue need to be investigated further in order to find the optimal preventive strategies and treatments.

References

1. Valstar E, Kaptein B, Nelissen R. Radiostereometry and new prostheses. Acta Orthop. Nordic Orthopaedic Federation Stockholm; 2012 Apr 24;83(2):103–4.
2. Evans AM, Rome K. A Cochrane review of the evidence for non-surgical interventions for flexible pediatric flat feet. Eur J Phys Rehabil Med. 2011;47(1):69–89.
3. Carr JB, Yang S, Lather LA. Pediatric Pes Planus: A State-of-the-Art Review. Pediatrics. 2016 Feb 17;137(3):1–10.
4. Mosca VS. Calcaneal lengthening for valgus deformity of the hindfoot. Results in children who had severe, symptomatic flatfoot and skewfoot. J Bone Joint Surg Am. The Journal of Bone and Joint Surgery, Inc.; 1995 Apr 1;77(4):500–12.
5. MacKenzie J, Rome K, Evans A. The Efficacy of Nonsurgical Interventions for Pediatric Flexible Flat Foot: A Critical Review. J Pediatr Orthop. 2012;32(8):830–4.
6. Roche AJ, Calder JDF. Lateral column lengthening osteotomies. Foot Ankle Clin. Elsevier; 2012 Jun;17(2):259–70.
7. Evans D. Calcaneo-valgus deformity. J Bone Joint Surg Br. 1975;57:270–8.
8. Mosca VS. Calcaneal Lengthening for Valgus Deformity of the Hindfoot. J Bone Jt Surg Am. 1995;
9. Blitz NM, Stabile RJ, Giorgini RJ, DiDomenico L a. Flexible Pediatric and Adolescent Pes Planovalgus: Conservative and Surgical Treatment Options. Clin Podiatr Med Surg. 2010;27:59–77.
10. Dumontier T a, Falicov A, Mosca V, Sangeorzan B. Calcaneal lengthening: investigation of deformity correction in a cadaver flatfoot model. Foot ankle Int / Am Orthop Foot Ankle Soc [and] Swiss Foot Ankle Soc. 2005;26(2):166–70.
11. Mosca VS. Principles and Management of Pediatric Foot and Ankle Deformities and Malformations. First. Brown B, Murphy D, editors. Philadelphia: Wolters Kluver Health; 2014. 1-275 p.
12. Mosca VS. Flexible flatfoot in children and adolescents. J Child Orthop. Springer; 2010 Apr 1;4(2):107–21.
13. Zeifang F, Breusch SJ, Döderlein L. Evans calcaneal lengthening procedure for spastic flexible flatfoot in 32 patients (46 feet) with a followup of 3 to 9 years. Foot ankle Int / Am Orthop Foot Ankle Soc [and] Swiss Foot Ankle Soc. 2006;27:500–7.
14. Adams SB, Simpson AW, Pugh LI, Stasikelis PJ. Calcaneocuboid joint subluxation after calcaneal lengthening for planovalgus foot deformity in children with cerebral

palsy. J Pediatr Orthop. 2009 Mar;29(2):170–4.
15. Dunn SP, Meyer J. Displacement of the anterior process of the calcaneus after evans calcaneal osteotomy. J Foot Ankle Surg. Elsevier Ltd; 2011;50(4):402–6.
16. Vining NC, Warme WJ, Mosca VS. Comparison of structural bone autografts and allografts in pediatric foot surgery. J Pediatr Orthop. 2012;32(7):719–23.
17. Prissel M a., Roukis TS. Incidence of Nonunion of the Unfixated, Isolated Evans Calcaneal Osteotomy: A Systematic Review. J Foot Ankle Surg. Elsevier Ltd; 2012;51(3):323–5.
18. Ahn J-Y, Lee H-S, Kim C-H, Yang J-P, Park S-S. Calcaneocuboid Joint Subluxation After the Calcaneal Lengthening Procedure in Children. Foot Ankle Int. 2014;35(March):1071100714528494.
19. Lee IH, Chung CY, Lee KM, Kwon S-S, Moon SY, Jung KJ, et al. Incidence and risk factors of allograft bone failure after calcaneal lengthening. Clin Orthop Relat Res. 2015 May;473(5):1765–74.
20. Zimmermann G, Moghaddam A. Allograft bone matrix versus synthetic bone graft substitutes. Injury. Elsevier Ltd; 2011;42:S16–21.
21. Giannoudis P V., Chris Arts JJ, Schmidmaier G, Larsson S. What should be the characteristics of the ideal bone graft substitute? Injury. Elsevier Ltd; 2011;42:S1–2.
22. Fillingham Y, Jacobs J. Bone grafts and their substitutes. Bone Joint J. 2016 Jan 1;98-B(1 Suppl A):6–9.
23. Banwart JC, Asher MA, Hassanein RS. Iliac crest bone graft harvest donor site morbidity. A statistical evaluation. Spine (Phila Pa 1976). 1995;20:1055–60.
24. Arrington ED, Smith WJ, Chambers HG, Bucknell AL, Davino NA. Complications of iliac crest bone graft harvesting. Clin Orthop Relat Res. 1996;300–9.
25. Kalk WW, Raghoebar GM, Jansma J, Boering G. Morbidity from iliac crest bone harvesting. J Oral Maxillofac Surg. 1996;54:1424–9; discussion 1430.
26. Seiler JG, Johnson J. Iliac crest autogenous bone grafting: donor site complications. J South Orthop Assoc. 2000 Jan;9(2):91–7.
27. Ahlmann E, Patzakis M, Roidis N, Shepherd L, Holtom P. Comparison of anterior and posterior iliac crest bone grafts in terms of harvest-site morbidity and functional outcomes. J Bone Joint Surg Am. 2002;84-A(May):716–20.
28. Shepherd L, Holtom P. Comparison of anterior and posterior iliac crest bone grafts in terms ... 2002;(May).
29. DeOrio JK, Farber DC. Morbidity associated with anterior iliac crest bone grafting in foot and ankle surgery. Foot ankle Int / Am Orthop Foot Ankle Soc [and] Swiss Foot Ankle Soc. 2005;26:147–51.
30. Sasso RC, LeHuec JC, Shaffrey C. Iliac crest bone graft donor site pain after anterior lumbar interbody fusion: a prospective patient satisfaction outcome assessment. J Spinal Disord Tech. 2005;18 Suppl:S77–81.
31. Singh JR, Nwosu U, Egol KA. Long-term functional outcome and donor-site morbidity

associated with autogenous iliac crest bone grafts utilizing a modified anterior approach. Bull NYU Hosp Jt Dis. 2009;67:347–51.

32. Schwartz CE, Martha JF, Kowalski P, Wang DA, Bode R, Li L, et al. Prospective evaluation of chronic pain associated with posterior autologous iliac crest bone graft harvest and its effect on postoperative outcome. Health Qual Life Outcomes. 2009;7:49.
33. Schaaf H, Lendeckel S, Howaldt HP, Streckbein P. Donor site morbidity after bone harvesting from the anterior iliac crest. Oral Surgery, Oral Med Oral Pathol Oral Radiol Endodontology. 2010;109:52–8.
34. Becker ST, Warnke PH, Behrens E, Wiltfang J. Morbidity after iliac crest bone graft harvesting over an anterior versus posterior approach. J Oral Maxillofac Surg. 2011;69:48–53.
35. Myeroff C, Archdeacon M. Autogenous Bone Graft : Donor Sites and Techniques. J Bone Jt Surg. Elsevier; 2011 Dec 7;93(23):2227–36.
36. Dimitriou R, Mataliotakis GI, Angoules AG, Kanakaris NK, Giannoudis P V. Complications following autologous bone graft harvesting from the iliac crest and using the RIA: A systematic review. Injury. Elsevier Ltd; 2011;42:S3–15.
37. Fasolis M, Boffano P, Ramieri G. Morbidity associated with anterior iliac crest bone graft. Oral Surg Oral Med Oral Pathol Oral Radiol. 2012;114:586–91.
38. Sung KH, Chung CY, Lee KM, Lee SY, Park MS. Calcaneal lengthening for planovalgus foot deformity in patients with cerebral palsy. Clin Orthop Relat Res. Springer; 2013 May 1;471(5):1682–90.
39. Yoo WJ, Chung CY, Choi IH, Cho T-J, Kim DH. Calcaneal lengthening for the planovalgus foot deformity in children with cerebral palsy. J Pediatr Orthop. 2005;25(6):781–5.
40. Kwak YH, Park KB, Park HW, Kim HW. Use of Allograft in Skeletally Immature Patients for Calcaneal Neck Lengthening Osteotomy. Yonsei Med J. Yonsei University College of Medicine; 2008;49(1):79–83.
41. Templin D, Jones K, Weiner DS. The Incorporation of Allogeneic and Autogenous Bone Graft in Healing of Lateral Column Lengthening of the Calcaneus. J Foot Ankle Surg. 2008;47(4):283–7.
42. Kadhim M, Holmes L, Miller F. Long-term Outcome of Planovalgus Foot Surgical Correction in Children with Cerebral Palsy. J Foot Ankle Surg. Elsevier Ltd; 2013;52(6):697–703.
43. Campana V, Milano G, Pagano E, Barba M, Cicione C, Salonna G, et al. Bone substitutes in orthopaedic surgery: from basic science to clinical practice. J Mater Sci Mater Med. Springer; 2014 Oct;25(10):2445–61.
44. Balakumar B, Babu S, Varma HK, Madhuri V. Triphasic ceramic scaffold in paediatric and adolescent bone defects. J Pediatr Orthop B. 2014;23:187–95.
45. Dolan CM, Henning J a., Anderson JG, Bohay DR, Kornmesser MJ, Endres TJ.

Randomized Prospective Study Comparing Tri-Cortical Iliac Crest Autograft to Allograft in the Lateral Column Lengthening Component for Operative Correction of Adult Acquired Flatfoot Deformity. Foot Ankle Int. 2007;28:8–12.

46. Lobo SE, Arinzeh TL. Biphasic calcium phosphate ceramics for bone regeneration and tissue engineering applications. Materials (Basel). 2010;3(2):815–26.
47. Kurien T, Pearson RG, Scammell BE. Bone graft substitutes currently available in orthopaedic practice: The evidence for their use. Bone Joint J. 2013 May;95-B(5):583–97.
48. Dorozhkin S V. Calcium orthophosphates as bioceramics: state of the art. J Funct Biomater. Multidisciplinary Digital Publishing Institute (MDPI); 2010 Jan 1;1(1):22–107.
49. Dorozhkin S V. Biocomposites and hybrid biomaterials based on calcium orthophosphates. Biomatter. Landes Bioscience; 2011 Jan 1;1(1):3–56.
50. Hannink G, Arts JJC. Bioresorbability, porosity and mechanical strength of bone substitutes: What is optimal for bone regeneration? Injury. Elsevier Ltd; 2011;42:S22–5.
51. Balçik C, Tokdemir T, Şenköylü A, Koç N, Timuçin M, Akin S, et al. Early weight bearing of porous HA/TCP (60/40) ceramics in vivo: A longitudinal study in a segmental bone defect model of rabbit. Acta Biomater. 2007;3:985–96.
52. Gouin F, Yaouanc F, Waast D, Melchior B, Delecrin J, Passuti N. Open wedge high tibial osteotomies: Calcium-phosphate ceramic spacer versus autologous bonegraft. Orthop Traumatol Surg Res. 2010;96:637–45.
53. Delécrin J, Takahashi S, Gouin F, Passuti N. A synthetic porous ceramic as a bone graft substitute in the surgical management of scoliosis: a prospective, randomized study. Spine. 2000.
54. Lindaman LM. Bone healing in children. Clin Podiatr Med Surg. 2001 Jan;18(1):97–108.
55. Lichte P, Pape HC, Pufe T, Kobbe P, Fischer H. Scaffolds for bone healing: Concepts, materials and evidence. Injury. 2011. p. 569–73.
56. Parithimarkalaignan S, Padmanabhan T V. Osseointegration: an update. J Indian Prosthodont Soc. Medknow Publications; 2013 Mar 1;13(1):2–6.
57. Anand A, Feldman DS, Patel RJ, Lehman WB, van Bosse HJP, Badra MI, et al. Interobserver and intraobserver reliability of radiographic evidence of bone healing at osteotomy sites. J Pediatr Orthop B. 2006;15(4):271–2.
58. Dayton P, Wienke JC, Prins DB, Haulard JP, Feilmeier MJ, Wiarda C. Measurement of Mid-Calcaneal Length on Plain Radiographs. Reliability of a New Method. Foot Ankle Spec. 2011;(October):280–3.
59. Valstar E, Kaptein B, Nelissen R. Radiostereometry and new prostheses. Acta Orthop. 2012 Apr;83(2):103–4.
60. Kärrholm J, Gill RHS, Valstar ER. The history and future of radiostereometric analysis. Clin Orthop Relat Res. 2006 Jul;448(448):10–21.
61. Nelissen RGHH. RSA and Registries: The Quest for Phased Introduction of New Implants. The Journal of Bone and Joint Surgery. 2011. p. 62.

62. Kärrholm J, Hansson L, Selvik G. Roentgen Stereophotogrammetric Analysis of Growth Pattern After Supination-Adduction Ankle Injuries in Children. J Pediatr Orthop. 1982;2(3).
63. Kärrholm J, Hansson LI, Selvik G. Longitudinal growth rate of the distal tibia and fibula in children. Clin Orthop Relat Res. 1984;(191):121–8.
64. Hildebrand H, Aronson S, Kullendorff CM, Selvik G. Roentgen stereophotogrammetric short-term analysis of growth rate in children operated for Crohn's disease. Acta Paediatr Scand. 1991;80(10):917–23.
65. Lauge-Pedersen H, Hägglund G, Johnsson R. Radiostereometric analysis for monitoring percutaneous physiodesis. J Bone Jt Surg Br Vol. 2006;88-B:1502–7.
66. Bragdon CR, Thanner J, Greene ME, Malchau H, Digas G, Harris WH, et al. Standing versus supine radiographs in RSA evaluation of femoral head penetration. Clin Orthop Relat Res. 2006;448(448):46–51.
67. Resch S, Ryd L, Stenström A, Yuan X. Measurement of the forefoot with roentgen stereophotogrammetry in hallux valgus surgery. Foot ankle Int / Am Orthop Foot Ankle Soc [and] Swiss Foot Ankle Soc. 1995;16:271–6.
68. Hansen TB, Stilling M. Equally good fixation of cemented and uncemented cups in total trapeziometacarpal joint prostheses. A randomized clinical RSA study with 2-year follow-up. Acta Orthop. 2013 Feb;84(1):98–105.
69. Lundberg A, Goldie I, Kalin B, Selvik G. Kinematics of the Ankle/Foot Complex: Plantarflexion and Dorsiflexion. Foot Ankle Int . 1989 Feb 1;9 (4):194–200.
70. Alberius P. Bone reactions to tantalum markers. A scanning electron microscopic study. Acta Anat (Basel). 1983;115(4):310–8.
71. Aronson AS, Jonsson N, Alberius P. Tantalum markers in radiography. An assessment of tissue reactions. Skeletal Radiol. 1985;14(3):207–11.
72. Ryd L, Yuan X, Löfgren H. Methods for determining the accuracy of radiostereometric analysis (RSA). Acta Orthop Scand. Informa UK Ltd UK; 2000 Aug 8;71(4):403–8.
73. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006 Jan;4:79.
74. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res. 2012;21(8):1305–14.
75. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.
76. Meadows KA. Patient-reported outcome measures: an overview. Br J Community Nurs. 2011;16(3):146–51.
77. Santana M-J, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. Qual Life Res. 2014;23(5):1505–13.
78. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures

on routine practice: a structured review. J Eval Clin Pract. 2006 Oct;12(5):559–68.
79. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ. 2015 Jan 10;350(feb10_14):g7818.
80. WHO. A practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment. Geneva; 2013.
81. Morris C, Doll HA, Wainwright A, Theologis T, Fitzpatrick R. The Oxford ankle foot questionnaire for children SCALING , RELIABILITY AND VALIDITY. 2008;90(11):1451–6.
82. Morris C, Doll H, Davies N, Wainwright A, Theologis T, Willett K, et al. The Oxford Ankle Foot Questionnaire for children: responsiveness and longitudinal validity. Qual Life Res. 2009 Dec;18(10):1367–76.
83. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. Jan;8(2):94–104.
84. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49.
85. Tavakol MRD. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–5.
86. Shrout PE, Fleiss JL. Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull. 1979;86(2):420–8.
87. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lance. 1986.
88. Kocher MS, Smith JT, Iversen MD, Brustowicz K, Ogunwole O, Andersen J, et al. Reliability, validity, and responsiveness of a modified International Knee Documentation Committee Subjective Knee Form (Pedi-IKDC) in children with knee disorders. Am J Sports Med. 2011 May;39(5):933–9.
89. Terwee CB, Bot SDM, de Boer MR, van der Windt D a WM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34–42.
90. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. The Standards for Reporting of Diagnostic Accuracy Group. Croat Med J. 2003;44(5):639–50.
91. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of Noninferiority. 2012;
92. Singh D. Nils Silfverskiöld (1888-1957) and gastrocnemius contracture. Foot Ankle Surg. 2013 Jun;19(2):135–8.
93. Nielsen S, Ruperto N, Herlin T, Pedersen FK. The Danish version of the Childhood

Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). Clin Exp Rheumatol. 2001;19(4 Suppl 23):S50–4.

94. Martinkevich P, Rahbek O, Moller-Madsen B, Soballe K, Stilling M. Precise and feasible measurements of lateral calcaneal lengthening osteotomies by radiostereometric analysis in cadaver feet. Bone Jt Res. 2015;4(5):78–83.
95. Valstar ER, Gill R, Ryd L, Flivik G, Börlin N, Kärrholm J. Guidelines for standardization of radiostereometry (RSA) of implants. Acta Orthop. 2005;76(4):563–72.
96. Söderkvist I, Wedin PA. Determining the movements of the skeleton using well-configured markers. J Biomech. Elsevier; 1993;26(12):1473–7.
97. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004 Aug;240(2):205–13.
98. Pagé MG, Katz J, Stinson J, Isaac L, Martin-Pichora AL, Campbell F. Validation of the numerical rating scale for pain intensity and unpleasantness in pediatric acute postoperative pain: sensitivity to change over time. J Pain. 2012 Apr;13(4):359–69.
99. Cavanagh PR, Rodgers MM. The arch index: A useful measure from footprints. J Biomech. 1987 Jan;20(5):547–51.
100. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10.
101. Ranstam J, Ryd L, Onsten I. Accurate accuracy assessment: review of basic principles. Acta Orthop Scand. 2000 Mar;71(1):106–8.
102. Hansen TB, Larsen K, Bjergelund L, Stilling M. Trapeziometacarpal joint implants can be evaluated by roentgen stereophotogrammetric analysis. J hand Surg Eur Vol. 2010;35(6):480–5.
103. Sealed Envelope Ltd 2012 [Internet]. [cited 2015 Nov 11]. Available from: www.sealedenvelope.com/power/continuous-noninferior
104. Davids JR, Gibson TW, Pugh LI. Quantitative segmental analysis of weight-bearing radiographs of the foot and ankle for children: normal alignment. - PubMed - NCBI. J Pediatr Orthop. 2005;25(6):769–76.
105. Martinkevich P, Møller-Madsen B, Gottliebsen M, Kjeldgaard Pedersen L, Rahbek O. Validation of the translated Oxford ankle foot questionnaire in 82 Danish children aged between five and 16 years. Bone Jt J. 2015;97B(3):420–6.
106. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord. 2010;11:144.
107. Kothari A, Stebbins J, Zavatsky AB, Theologis T. Health-related quality of life in children with flexible flatfeet: a cross-sectional study. J Child Orthop. 2014 Dec;8(6):489–96.
108. Eiser C, Varni JW. Health-related quality of life and symptom reporting: Similarities

and differences between children and their parents. Eur J Pediatr. 2013;172(10):1299–304.

109. Martinelli N, Romeo G, Bonifacini C, Viganò M, Bianchi A, Malerba F. Validation of the Italian version of the Oxford Ankle Foot Questionnaire for children. Qual Life Res. 2016 Jan;25(1):117–23.
110. Glazebrook M, Younger A, Wing K, Lalonde K-A. A prospective pilot study of B2A-coated ceramic granules (Amplex) compared to autograft for ankle and hindfoot arthrodesis. Foot Ankle Int. 2013 Aug;34(8):1055–63.
111. Morshed S. Current Options for Determining Fracture Union. Adv Med. 2014 Jan;2014:708574.
112. Ploghaus A, Narain C, Beckmann CF, Clare S, Bantick S, Wise R, et al. Exacerbation of pain by anxiety is associated with activity in a hippocampal network. J Neurosci. 2001 Dec 15;21(24):9896–903.
113. Clarke A, Flowers MJ, Davies AG, Fernandes J, Jones S. Morbidity associated with anterior iliac crest bone graft harvesting in children undergoing orthopaedic surgery: a prospective review. J Child Orthop. Springer Berlin Heidelberg; 2015;9(5):411–6.
114. Zimmerli W. Clinical presentation and treatment of orthopaedic implant-associated infection. J Intern Med. 2014 Aug;276(2):111–9.
115. Becker K, Heilmann C, Peters G. Coagulase-Negative Staphylococci. Clin Microbiol Rev. 2014 Oct 2;27(4):870–926.
116. Borgers N, Leeuw E De, Hox J. Children as Respondents in Survey Research: Cognitive Development and Response Quality. Bull Methodol Sociol. 2000;66(April):60–75.
117. Rabello F, Souza C, Farina Jr J. Update on hypertrophic scar treatment. Clinics. Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo; 2014 Aug 12;69(8):565–73.
118. Scholl R, Bekker A, Babu R. Neuroendocrine and Immune Responses to Surgery. Internet J Anesthesiol. 2012;30(3).
119. Kozlowski LJ, Kost-Byerly S, Colantuoni E, Thompson CB, Vasquenza KJ, Rothman SK, et al. Pain prevalence, intensity, assessment and management in a hospitalized pediatric population. Pain Manag Nurs. 2014 Mar;15(1):22–35.
120. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. BioMed Central Ltd; 2011 Jan 3;12(1):106.
121. Campbell AJ, Bagley A, Van Heest A, James MA. Challenges of randomized controlled surgical trials. Orthop Clin North Am. 2010 Apr;41(2):145–55.
122. Van Heest AE, Bagley A, Molitor F, James MA. Tendon Transfer Surgery in Upper-Extremity Cerebral Palsy Is More Effective Than Botulinum Toxin Injection or Regular, Ongoing Therapy. J Bone Jt Surg. Elsevier; 2015;97-A(7):529–36.
123. Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. Health Technol Assess. 2005

Mar;9(7):1–238, iii – iv.

124. Ware JH, Muller JE, Braunwald E. The futility index. An approach to the cost-effective termination of randomized clinical trials. Am J Med. 1985 Apr;78(4):635–43.
125. Pocock SJ. When to stop a clinical trial. BMJ. 1992 Jul 25;305(6847):235–40.
126. Müller S, Carlsohn A, Müller J, Baur H, Mayer F. Static and dynamic foot characteristics in children aged 1-13 years: A cross-sectional study. Gait Posture. 2012;35(3):389–94.
127. Kothari A, Dixon PC, Stebbins J, Zavatsky AB, Theologis T. The relationship between quality of life and foot function in children with flexible flatfeet. Gait Posture. 2015 Mar;41(3):786–90.
128. Kothari A, Dixon PC, Stebbins J, Zavatsky AB, Theologis T. The relationship between quality of life and foot function in children with flexible flatfeet. Gait Posture. 2015 Mar;41(3):786–90.
129. Wang B, Roach KE, Kapron AL, Fiorentino NM, Saltzman CL, Singer M, et al. Gait & Posture Accuracy and feasibility of high-speed dual fluoroscopy and model-based tracking to measure in vivo ankle arthrokinematics. Gait Posture. Elsevier B.V.; 2015;41(4):888–93.
130. Burrows M, Liu D, McKay H. High-resolution peripheral QCT imaging of bone micro-structure in adolescents. Osteoporis Int. 2010;21(3):515–20.
131. Mueller MJ, Strube MJ. Generalizability of in-shoe peak pressure measures using the F-scan system. Clin Biomech (Bristol, Avon). 1996 Apr;11(3):159–64.
132. Zammit G V, Menz HB, Munteanu SE. Reliability of the TekScan MatScan(R) system for the measurement of plantar forces and pressures during barefoot level walking in healthy adults. J Foot Ankle Res. 2010;3:11.

Appendices

Appendix 1

Dynamic plantar pressure assessment protocol

The patients were instructed to walk back and forth at a self-selected pace, as naturally as possible crossing the platform without targeting it. The walking distance was three meters with the platform placed in the middle. As the patients were familiarised with the procedure, the recordings were initiated in accordance with a midgait protocol. A minimum of three passes of the foot of interest were collected as this has been shown to be sufficient for attaining high reliability. [131,132] The right foot must strike the pressure plate from the right side and the left foot must strike the pressure plate from the left side.

Step calibration with subject body mass was performed at each inital examination.

The pressure mapping system used was of piezo-resistive technology, the high-resolution mat system, HR Mat™ (Tekscan® Inc. South Boston, MA, USA). The sensing area measured 48.77 x 44.70 cm with a sensor spatial resolution of 4 sensels/cm^2, and a total of 8448 sensing elements. A sampling frequency of 66.6 Hz was used in the study. The analysis is in this thesis is restricted to the arch index assessed as the contact area of the midfoot relative to the total contact area of the plantar surface area exlcuding the toes.

Data were analysed with the F-Scan Research version 6.70-03 software. Each movie was reviewed and insufficient foot steps (partial, or striking the platformed from the wrong direction) were excluded. The peak stance averaging option was chosen for analysis, inlcuding only complete foot steps.

The research software allows for adjusting the automated foot masks for each individual. Twelve mask are generated: The medial heel, lateral heel, midfoot, metatarsal 1 to 5, toe 1 to 3 and 4th and 5th toes together.

The quantitative part comprised calculating the relative contact area and peak pressure of each mask as a fraction of the total mask sum, thus enabling comparison between patients of different body mass and plantar surface area.

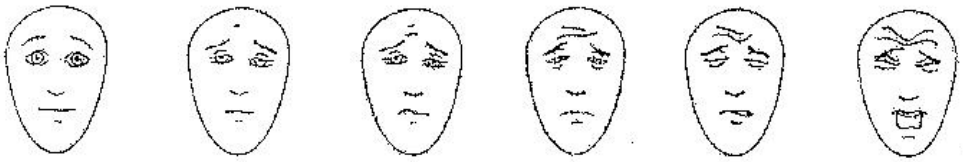
Appendix 2

Pain assessment sheet

Patients were instructed to the sheet and asked to complete it at given time points.

Der findes forskellige skalaer til vurdering af smerteintensitet

Bieris ansigtsskala



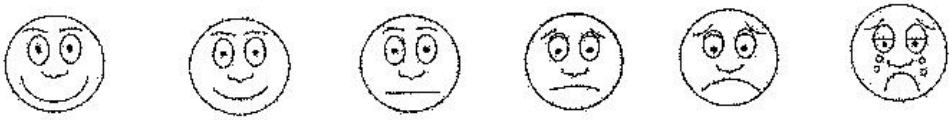
Numerisk Rang Skala

0 1 2 3 4 5 6 7 8 9 10

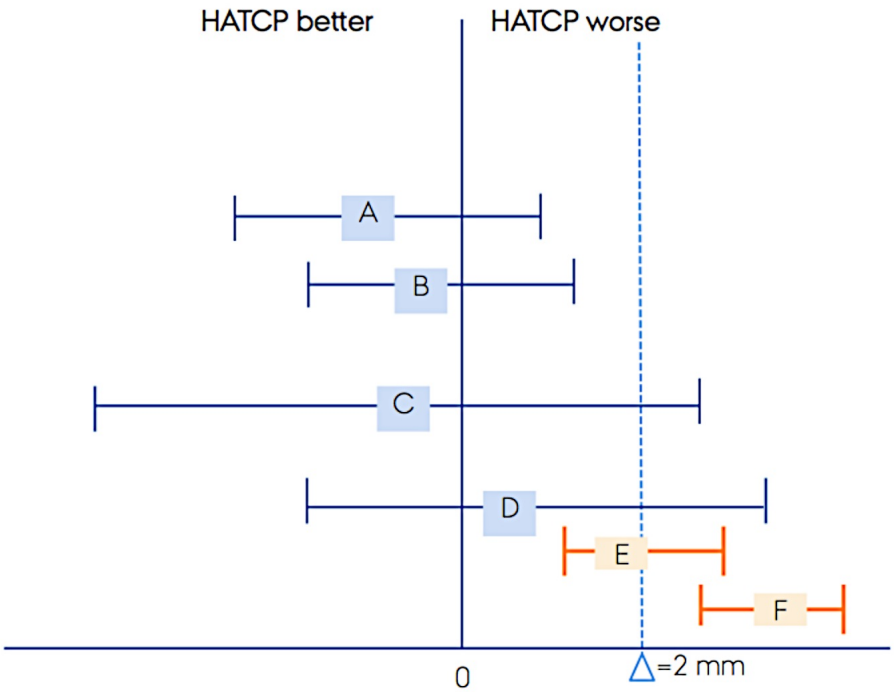
Visuel Analog Skala

Ingen smerter **Værst tænkelige smerter**

Wong Baker ansigtsskala



Appendix 3



Non-inferiority figure. Adapted from Piaggio et al 2012[91] in the CONSORT guidelines. The figure has been adjusted to fit the interim-analysis of this thesis.
The error bars indicate 2-sided 99.8% CI, with each lettter describing possible outcomes in a non-inferiority study. The difference in means is calculated by μHATCP(Tx)- μAUTO(Tx). The line at zero indicates no difference between the two treatments. The dashed line at Δ is the non-inferiority margin with Δ = 2 mm. NB. In *scenario A* and *B*, the CI includes zero and lies to the left of the Δ. This means, the new treatment is non-inferior but is has not been shown to be superior. *Scenario C* and *D* include zero and Δ. The difference is not significant and inconclusive regarding non-inferiority. *Scenario E lies to the right of zero and includes* Δ, whice means the new treatnent is significantly worse, but the results are inconclusive regarding magnitude of possible inferiority. *Scenario F* represents inferiority of the new treatment. If *scenario E* or *F* were encountered in *Study III*, the study would be stopped.

Appendix 4

Oxford Ankle Foot Questionnaire item scores

OxAFQ-children item median (range) scores at baseline, 6 and 12 months follow-up from the perspective of children in the **AUTO** group.

AUTO			
Item	Baseline (n=5)	6 months follow-up (n=5)	12 months follow-up (n=3)
	median (range)		
Q1	2 (2-3)	4 (2-4)	4 (4-4)
Q2	2 (0-3)	3 (1-4)	3 (2-4)
Q3	3 (2-4)	4 (2-4)	4 (4-4)
Q4	3 (2-4)	3 (1-4)	4 (4-4)
Q5	3 (1-4)	4 (2-4)	4 (3-4)
Q6	2 (2-4)	3 (2-4)	4 (3-4)
Q7	4 (2-4)	4 (2-4)	4 (4-4)
Q8	4 (2-4)	4 (2-4)	4 (4-4)
Q9	3 (1-4)	4 (2-4)	4 (2-4)
Q10	4 (3-4)	4 (3-4)	4 (4-4)
Q11	4 (1-4)	4 (1-4)	4 (1-4)
Q12	4 (3-4)	3 (2-4)	3 (1-4)
Q13	4 (3-4)	4 (2-4)	4 (1-4)
Q14	4 (4-4)	4 (4-4)	4 (4-4)
Q15	2 (1-3)	4 (1-4)	3 (0-4)

OxAFQ-children item median (range) scores at baseline, 6 and 12 months follow-up from the perspective of all children in the **HATCP** group (including the reoperated patient's baseline scores). Two of the six patients were not cabaple of completing the questionnaire.

HATCP			
Item	Baseline (n=4)	6 months follow-up (n=3)	12 months follow-up (n=3)
	median (range)		
Q1	2 (0-2)	2 (2-3)	3 (2-3)
Q2	1(0-1)	2 (1-2)	2 (1-3)
Q3	2 (0-2)	3 (2-3)	2 (2-2)
Q4	1 (0-2)	2 (2-2)	2 (2-3)
Q5	1 (0-4)	2 (2-2)	2 (1-2)
Q6	2 (2-2)	3 (2-3)	3 (3-3)
Q7	1 (1-2)	3 (2-4)	3 (2-4)
Q8	1 (1-2)	3 (3-3)	3 (2-3)
Q9	1 (0-2)	0 (0-2)	0 (0-3)
Q10	3 (2-4)	4 (3-4)	2 (2-4)
Q11	3 (2-4)	4 (2-4)	3 (3-4)
Q12	1 (0-2)	3 (3-4)	4 (3-4)
Q13	3 (2-4)	4 (4-4)	4 (4-4)
Q14	4 (2-4)	4 (4-4)	4 (4-4)
Q15	1 (0-2)	2 (1-2)	2 (1-3)

OxAFQ-parents item median (range) scores at baseline, 6 and months follow-up from the perspective of the parents.

AUTO			
Item	Baseline (n=5)	6 months follow-up (n=5)	12 months follow-up (n=3)
	median (range)		
Q1	2 (1-4)	4 (1-4)	4 (3-4)
Q2	2 (0-4)	3 (0-4)	4 (2-4)
Q3	3 (1-4)	4 (2-4)	4 (2-4)
Q4	1 (1-2)	4 (1-4)	4 (4-4)
Q5	2 (0-3)	3 (3-4)	4 (3-4)
Q6	1 (0-3)	3 (2-4)	4 (1-4)
Q7	3 (2-4)	4 (2-4)	4 (4-4)
Q8	3 (2-4)	4 (1-4)	4 (3-4)
Q9	3 (2-4)	4 (1-4)	4 (2-4)
Q10	3 (3-4)	4 (2-4)	4 (3-4)
Q11	3 (2-4)	4 (1-4)	4 (0-4)
Q12	3 (1-4)	3 (1-4)	4 (0-4)
Q13	4 (2-4)	4 (2-4)	4 (1-4)
Q14	4 (2-4)	4 (2-4)	4 (2-4)
Q15	2 (0-3)	4 (0-4)	4 (0-4)

OxAFQ-p item mean scores at baseline, 6 and 12 months follow-up from the perspective of the parents in the HATCP group. Including the parent to the reoperated patient.

HATCP			
Item	Baseline (n=6)	6 months follow-up (n=5)	12 months follow-up (n=5)
	median (range)		
Q1	2 (1-3)	2 (2-4)	3 (2-4)
Q2	1 (0-1)	2 (1-3)	3 (1-4)
Q3	2 (1-4)	2 (2-4)	3 (2-3)
Q4	1 (1-2)	2 (1-4)	3 (2-4)
Q5	1 (0-1)	2 (0-3)	3 (1-4)
Q6	2 (1-2)	3 (1-4)	3 (3-4)
Q7	2 (1-3)	2 (2-4)	3 (1-4)
Q8	2 (1-3)	3 (2-4)	3 (2-4)
Q9	1 (0-2)	2 (0-4)	2 (0-4)
Q10	3 (2-4)	4 (3-4)	4 (3-4)
Q11	4 (2-4)	4 (3-4)	4 (4-4)
Q12	2 (1-4)	4 (2-4)	3 (2-4)
Q13	3 (2-4)	4 (2-4)	4 (1-4)
Q14	3 (1-4)	4 (4-4)	4 (3-4)
Q15	1 (0-1)	2 (0-4)	3 (2-4)

Appendix 5

Oxford Ankle Foot Questionnaire - original English version

The Oxford Ankle Foot Questionnaire - Child version (the parent (proxy) version is rephrased by using the prefix "Has your child...")

The questions below are based upon ways in which some young people told us they had been affected by a foot or ankle problem. We want you to think about each question and then put a tick or a cross next to the answer that best describes you – was it never a problem for you, or was it always a problem for you, or somewhere in between?

Thinking about last week...

Q1 Have you found walking difficult because of your foot and ankle?

Never Rarely Sometimes Very often Always

Q2 Have you found it difficult to run because of your foot and ankle?

Never Rarely Sometimes Very often Always

Q3 Has it been difficult to stand up for long periods?

Never Rarely Sometimes Very often Always

Q4 Have you had pain in your foot and ankle?

Never Rarely Sometimes Very often Always

Q5 Have your legs been sore or ached after walking or running?

Never Rarely Sometimes Very often Always

Q6 Have you felt tired because of your foot and ankle?

Never Rarely Sometimes Very often Always

Q7 Has your foot or ankle stopped you joining in with other in the playground?

Never Rarely Sometimes Very often Always

Q8 Has your foot or ankle stopped you playing in the park or outside?

Never Rarely Sometimes Very often Always

Q9 Has your foot or ankle stopped you in taking part in PE lessons?

Never Rarely Sometimes Very often Always

Q10 Has your foot or ankle stopped you in taking part in any other lessons at school?

Never Rarely Sometimes Very often Always

Q11 Have you been bothered by how your foot or ankle looks?

Never Rarely Sometimes Very often Always

Q12 Has the way you walked bothered you?

Never Rarely Sometimes Very often Always

Q13 Have you been embarrassed because of your foot or ankle?

Never Rarely Sometimes Very often Always

Q14 Has anyone been unkind to you because of your foot or ankle?

Never Rarely Sometimes Very often Always

Q15 Has your foot or ankle stopped you wearing any shoes you wanted to wear?

Never Rarely Sometimes Very often Always

Paper I

■ CHILDREN'S ORTHOPAEDICS

Validation of the translated Oxford ankle foot questionnaire in 82 Danish children aged between five and 16 years

P. Martinkevich,
B. Møller-Madsen,
M. Gottliebsen,
L. Kjeldgaard Pedersen,
O. Rahbek

From Department of Children's Orthopaedics, Aarhus University Hospital, Denmark

We present the validation of a translation into Danish of the Oxford ankle foot questionnaire (OxAFQ). We followed the Isis Pros guidelines for translation and pilot-tested the questionnaire on ten children and their parents. Following modifications we tested the validity of the final questionnaire on 82 children (36 boys and 45 girls) with a mean age of 11.7 years (5.5 to 16.0) and their parents. We tested the reliability (repeatability (test–retest), child–parent agreement, internal consistency), feasibility (response rate, time to completion, floor and ceiling effects) and construct validity. The generic child health questionnaire was used for comparison. We found good internal consistency for the physical and the school and play domains, but lower internal consistency for the emotional domain. Overall, good repeatability was found within children and parents as well as agreement between children and parents. The OxAFQ was fast and easy to complete, but we observed a tendency towards ceiling effects in the school and play and emotional domains. To our knowledge this is the first independent validation of the OxAFQ in any language. We found it valid and feasible for use in the clinic to assess the impact on children's lives of foot and/or ankle disorders. It is a valuable research tool.

Cite this article: *Bone Joint J* 2015;97-B:420–6.

The Oxford ankle foot questionnaire (OxAFQ) for children was first described in 2008[1] to assess disability associated with foot and ankle problems in children aged between five and 16 years. It is a region-specific health-related patient-reported outcome measure (PROM) consisting of 14 questions providing face validity. It covers three domains as experienced by the child and his/her parents during the previous week: physical activity (six items), school and play (four items), emotional (four items) and an item relating to footwear.[1] There are four possible answers for each item, scoring between zero and four points, with four representing the best outcome.

The OxAFQ has been validated as a discriminative and evaluative tool that can differentiate between severe chronic disease and benign fluctuating disease.[2] It can also detect improvement or deterioration in symptoms, sports activity and emotional wellbeing in children with a variety of ankle and foot disorders.[1,2] It has been found by its developers to have cross-sectional validity, internal test–retest reliability, responsiveness and longitudinal validity.[1,2] Two versions are available: OxAFQ-c for five- to 16-year-olds and a parent-proxy version, OxAFQ-p. For simplicity, both versions are referred to here as OxAFQ.

The child health questionnaire (CHQ) is a generic outcome instrument designed to assess health-related quality of life in children and adolescents independently of their underlying diagnosis.[3] The psychometric properties of the CHQ have been thoroughly validated in Danish.[4] We selected the CHQ as a comparison tool because of the lack of related validated PROMs when we began our study.

This study comprised two stages. In the first, cross-cultural validation stage, our aim was to perform a workable translation and cross-cultural adaptation of the English version of OxAFQ into Danish. In the second, the validation process, our aim was to test the agreement and reliability (internal consistency, test–retest reliability), feasibility (response rate, time to completion, floor and ceiling effects) and construct validity (hypothesis testing) of both the child and the parent Danish versions. To our knowledge, the OxAFQ has not been validated by any group other than the developers. The translation of the original English version of OxAFQ adhered strictly to Isis Pros guidelines,[5] which are in line with the ISPOR Task Force Translation guidelines.[6] The translation from English to Danish was performed independently by two medical doctors (LKP, MG), who were both native Danish and who were

■ P. Martinkevich, MD, PhD student, Department of Children's Orthopaedics
■ B. Møller-Madsen, MD, DMSci, Professor, Department of Children's Orthopaedics
■ M. Gottliebsen, MD, PhD, Orthopaedic Surgeon, Department of Children's Orthopaedics
■ L. Kjeldgaard Pedersen, MD, PhD student, Department of Children's Orthopaedics
■ O. Rahbek, MD, PhD, Associate Professor, Department of Children's Orthopaedics
Aarhus University Hospital, Noerrebrogade 44, 8000 Aarhus C, Denmark.

Correspondence should be sent to Dr P. Martinkevich; e-mail: polina.martinkevich@gmail.com

doi:10.1302/0301-620X.97B3.34355 $2.00

Bone Joint J 2015;95-B:420–6.
Received 19 June 2014; Accepted after revision 26 November 2014

also fluent in English. The translations were reviewed by the principal investigator (PM), who produced a reconciled version.

The reconciled Danish version was translated back into English by two independent translators (KS, ES) who were fluent in both English and Danish. The translators did not see the original English version. A comprehensive review of this translation against the original version was conducted by three investigators (MG, LKP, PM) in order to make revisions and eliminate discrepancies.

Pilot testing

The initial translation was tested on ten Danish children and their parents attending an orthopaedic outpatient clinic. The children were a heterogeneous group suffering from various foot and ankle conditions. A healthcare professional administered the questionnaires.

Owing to cultural differences, all participants had difficulties understanding item 8 which was therefore modified to apply to Danish culture. All found item 4 odd but comprehensible. Furthermore, a few words that were understood by parents but not by children were simplified.

The project manager and the developer conducted the pilot-testing review. The queries were forwarded to the principal investigator and consensus on the best possible translation was reached.

Two sets of proofs were read, one by the principal investigator and the other by a translator who had not previously been involved in the translation process. No changes were recommended.

All translators met the criteria of the Isis PROs guidelines.

Validation process

Patients and methods

The study had approval from the Central Denmark Region Committee on Biomedical Research Ethics (Jr.nr. 1-10-72-250-12). Informed consent was obtained prior to inclusion in the study. This was conducted as a prospective observational study in the children's orthopaedic clinic between February 2013 and January 2014. Children were eligible for inclusion if they were understood the Danish language, were aged between five and 16 years, and were attending the elective outpatient clinic because of a foot and/or ankle disorder. Children were excluded if they had any significant proximal musculoskeletal symptoms. If a trauma should occur within a recent time frame, patients would stil be included.

Children and parents were invited to participate by the investigator (PM) or an assistant (GS) in the clinic prior to consultation with an orthopaedic surgeon. Children were divided into three groups: those who were newly referred from a general practitioner or another hospital (group 0), a clinical control group of those who had already been treated i.e. conservative or surgical interventions (group 1) and pre-admission children awaiting surgery (group 2).

At inclusion, participants were given the OxAF and CH questionnaires. The child received the CHQ-CF87 (long version; no short version is available for children), and the parents received the CHQ-PF28 (short version) to be completed with regard to their child. Children aged < nine years were not asked to complete the CHQ-CF87, as we took advice from one of the questionnaire' validators, who suggested that this age group has difficulties understanding the questionnaire. When necessary, parents were allowed to read the questionnaires out loud to their child, but without interpreting the questions.

Sample size considerations. Sample size calculations for validation studies of PROMs depend on the clinical purpose and/or the desired statistical test(s). In this study we assessed the measurement error properties of the OxAFQ using the Bland–Altman statistical methods as the primary test of interest.

PROMs are expected to be less accurate than quantitative methods of measurement, therefore we accepted an accuracy of 0.4 × standard deviation (SD) based on evaluations of the developers' original OxAFQ data[2] and recommendations by Bland.[7] This accuracy required a sample size of about 70 children, which is also an acceptable number of patients for a Bland–Altman plot and the other planned statistical methods. We expected to be able to recruit this number over the intended study period of one year.

All participants completed the Danish versions of both questionnaires in the clinic. They all received a second OxAFQ which they were asked to complete at home, two days after the first questionnaire, for test–retest reliability. On the second day the parents received a reminder text message and/or a phone call. Both administrations were independent, i.e. the participants did not know the scores from the initial questionnaire. None of the children were expected to have any changes in their condition during these two days.

Internal consistency. This reflects whether items in a scale measure the same underlying construct. Each domain of the OxAFQ is based on a reflective model, i.e., the items within each domain should be correlated, measuring the same construct. The initial questionnaire was used to assess internal consistency using Cronbach's alpha statistics. A score of between 0.7 and 0.9 reflects acceptable to good consistency.[8] This parameter was included for comparison with the original paper by Morris et al,[2] but as this is a difficult parameter to interpret we did not perform sample size calculations.

Reliability (relative measurement error). This describes the extent to which participants can be distinguished from each other despite errors of measurement.[9] It reflects the relative error, assessed as the intraclass correlation coefficient (ICC) both within (test–retest) and between participants (child–parent reliability). ICC based on a one-way random effects model (ICC 1.1) was used in this study.[10] Although we calculated the ICC for comparison with the developer's ICC it is clear that direct comparison is difficult, as the

Table I. Diagnosis at the time of inclusion in the study

Diagnosis (total n = 85*)	n
Pes planovalgus/pes planus	22
Congenital talipes equinovarus	10
Idiopathic toe-walker	8
Apophysitis/tendinitis	8
Macrodactylia	5
Minor deformities†	12
Trauma‡	7
Other§	13

* Some had more than one diagnosis
† Curly toes, hallux valgus, metatarsal deformities, os naviculare accessorium, syndactyly
‡ Trauma sequelae
§ Multiple hereditary exostoses, ingrown infected nails, osteoid osteoma, restricted ankle range of movement, drop foot, short lateral malleolus, tarsal coalition and non-specified pain

developers did not describe the type of ICC which they used. For this method, it is difficult to estimate the width of the confidence interval (CI), but at least 30 participants are recommended by Bonnet et al.[11]

Absolute error of measurement (repeatability and agreement). This can be defined as the degree to which the scores on repeated measures are close to each other.[9] It is assessed as limits of agreement (LOA) according to Bland and Altman[12] within subjects (retest responders) and between children and parents (all participants) for each domain and the footwear item. As the retest was completed at home, a control of the child–parent agreement between both data sets was performed to identify any bias using paired *t*-tests and Bland–Altman plots. A low limit of agreement is indicated by a small interval between the results, and low bias by a mean difference close to zero.

Construct validity. This is defined by the degree to which scores on a particular PROM are consistent with hypotheses, and is based on the assumption that the PROM validly measures the construct to be assessed.[13] Discriminant validity in terms of physical properties was assessed by the hypothesis that the pre-admission patients (group 2) would have lower scores in the physical domain than the clinical control group (group 1). The internal relationships of patient groups were illustrated by box plots. Convergent validity was assessed by eight hypotheses on the expected correlations of the OxAFQ domains with relevant CHQ domains using Spearman rank correlation, for which a sample size of at least seven observations is required. We took a conservative two-tailed approach, with significance indicated by $p = 0.05$. Based on earlier literature a moderate correlation ($rho > 0.4$)[14] was assumed between the following domains in the OxAFQ and the CHQ:

- The physical domain with physical functioning;
- The physical domain with role/social limitations – physical;
- The physical domain with bodily pain;
- The school and play domains with physical functioning;
- The school and play domains with role/social limitations – physical;
- The school and play domains with bodily pain;
- The emotional domain with mental health;
- The emotional domain with self-esteem.

Feasibility. This was assessed by the rate of response (missing responses) at inclusion, time to completeness, and floor and ceiling effects. The floor and ceiling effects were considered acceptable if < 30%.[15]

Statistical analysis. All statistical analyses were performed in Stata IC 12.1 for Macintosh, (StataCorp, Texas, USA). All tests were two-tailed, with a power of 0.9 and statistical significance assumed at $p < 0.05$.

Domain scale scores were calculated as the total score divided by the number of items in the particular domain. Missing data in the OxAFQ were handled in line with the developers' approach.[1] If half or more of the items in a domain scale were present, the missing data were imputed by the mean of the participant's item scores in that scale. Any missing CHQ data were handled according to the developers' manual for the individual domains. Domain scale scores were treated as continuous data. Data were tested for normality by histograms and QQ-plots, and assumptions of equal variances were checked prior to conducting the parametrical analysis.

Results

A total of 82 children and 81 parents were included in the study There were 36 boys and 46 girls, with a mean age of 11.7 years (5.5 to 16), which followed a normal distribution as examined by a QQ-plot. The retest responders comprised 24 boys and 30 girls, with a mean age of 11.5 years (5.5 to 16). No significant differences were found in age or gender between the responders in the initial test and the retest ($p = 0.49$, chi-squared test).

The children had a variety of foot and ankle disorders (Table I). The mean domain scores at the first test and retest for all participants are shown in Table II.

Internal consistency. Crohnbach's alpha indicated good internal consistency for the items in the physical and school and play domains, and moderate internal consistency for the emotional domain (Table III).

Reliability and measurement error (agreement). The mean retest response time was two days (zero to nine). A total of

Table II. Mean (standard deviation (SD)) and range baseline and retest domain scores in children and parents

	Baseline			Retest		
	n	Mean (SD)	Range	n	Mean (SD)	Range
Physical						
Child	82	2.5 (0.9)	0.5 to 4	54	2.7 (0.9)	0.3 to 4
Parent	81	2.4 (1.0)	0 to 4	53	2.7 (0.9)	0.9 to 4
School and play						
Child	82	3.1 (1.0)	0.3 to 4	54	3.2 (0.9)	1 to 4
Parent	81	3.1 (0.9)	0.8 to 4	53	3.3 (0.9)	1 to 4
Emotional						
Child	82	3.3 (0.7)	0.3 to 4	54	3.4 (0.7)	1 to 4
Parent	81	3.1 (0.8)	0.5 to 4	53	3.4 (0.6)	1.8 to 4
Footwear item						
Child	82	2.3 (1.5)	0 to 4	54	2.3 (1.3)	0 to 4
Parent	81	2.1 (1.4)	0 to 4	53	2.3 (1.1)	0 to 4

Table III. Crohnbach's alpha of the Oxford ankle foot questionnaire domains

	Physical	School and play	Emotional
Child (n = 82)	0.87	0.85	0.67
Parent (n = 81)	0.88	0.87	0.75

Table IV. Reliability coefficient and agreement for Oxford ankle foot questionnaire-c test–retest (n = 54)

	$Mean_{diff}$* (SD_{diff})	95% CI†	LOA‡	ICC§	95% CI¶
Physical	-0.12 (0.57)	-0.28 to 0.03	-1.27 to 1.02	0.81	0.72 to 0.89
School and play	0.005 (0.45)	-0.12 to 0.13	-0.88 to 0.89	0.84	0.77 to 0.92
Emotional	0.005 (0.63)	-0.17 to 0.18	-1.26 to 1.27	0.64	0.48 to 0.79
Footwear item	0.15 (0.92)	-0.10 to 0.40	-1.69 to 1.99	0.79	0.69 to 0.88

* Difference in means between first and second tests
† 95% confidence interval (CI) for difference in means
‡ 95% limits of agreement (LOA)
§ Intraclass correlation coefficient (ICC)
¶ 95% CI for intraclass correlation coefficient
SD, standard deviation

Table V. Reliability coefficient and agreement for Oxford ankle foot questionnaire-p test–retest (n = 53)

	$Mean_{diff}$ (SD_{diff})	95% CI	LOA	ICC	95% CI
Physical	-0.13 (0.56)	-0.28 to 0.03	-1.25 to 1.00	0.82	0.74 to 0.90
School and play	-0.12 (0.46)	-0.25 to 0.003	-1.04 to 0.79	0.87	0.81 to 0.93
Emotional	-0.09 (0.51)	-0.23 to 0.05	-1.10 to 0.92	0.76	0.65 to 0.87
Footwear item	0.06 (1.05)	-0.23 to 0.34	-1.69 to 1.99	0.69	0.55 to 0.82

LOA, limits of agreement; ICC, intraclass correlation coefficient; CI, confidence interval; SD, standard deviation

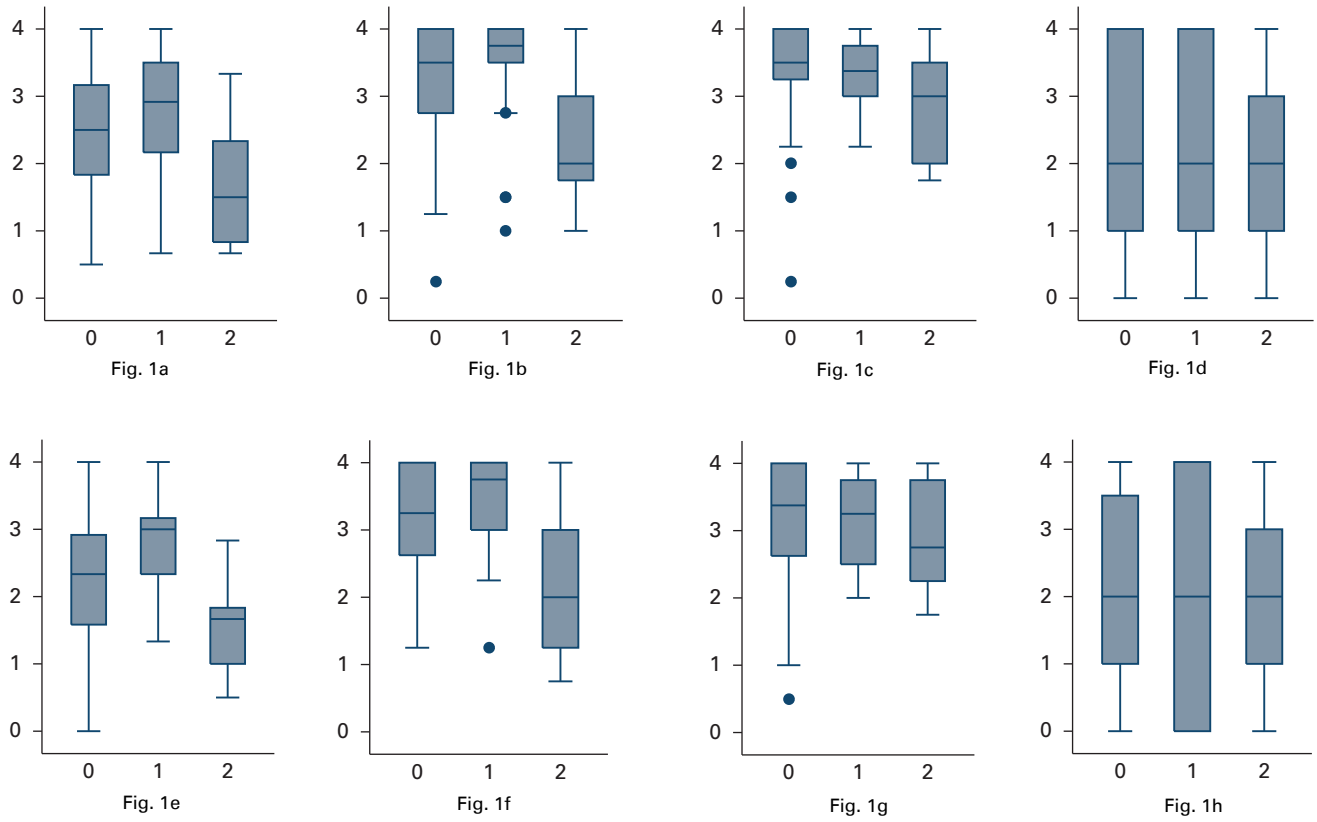
54 children and 53 parents responded to the retest (Table II), a response rate of approximately 65%. Tables IV and V show the reliability coefficients, systematic and random variations and LOA for test–retests and child–parent tests. ICC was good overall for all three domains. The absolute measurement errors within and between participant properties, judged by the SD_{diff}, were acceptable, with values between 0.45 and 0.77. The mean differences were small and revealed no systematic bias. The limits of agreement showed acceptable repeatability within both children (Table IV) and parents (Table V) for the three domains, with a maximum range between -1.26 and 1.27 (emotional domain). The agreement between children and parents was slightly broader (Table VI). The footwear item showed high random variation, SD_{diff} and broad limits of agreement.

Construct validity. There were 49 children in group 0.22 in group 1 and 11 in group 2. The box plots (Fig. 1) illustrate the mean scores in each domain, and the raw score in the footwear item is plotted against each study group for both children and parents. The hypothesis that the

Table VI. Reliability coefficient and agreement between Oxford ankle foot questionnaire-c and Oxford ankle foot questionnaire-p (n = 81)

	$Mean_{diff}$ (SD_{diff})	95% CI	LOA	ICC	95% CI
Physical	0.12 (0.77)	–0.05 to 0.29	–1.42 to 1.65	0.68	0.56 to 0.80
School and play	0.06 (0.75)	–0.11 to 0.22	–1.45 to 1.56	0.69	0.57 to 0.80
Emotional	0.20 (0.75)	0.03 to 036	–1.30 to 1.69	0.51	0.34 to 0.67
Foot wear item	0.21 (1.51)	–0.13 to 0.55	–2.82 to 3.24	0.45	0.27 to 0.62

SD, standard deviation; CI, confidence interval; LOA, limits of agreement; ICC, intraclass correlation coefficient



Box plots showing mean domain scores and footwear item raw scores within each study cohort: a) physical – child, b) school and play – child, c) emotional – child, d) footwear item – child; e) physical – parents, f) school and play – parents, g) emotional – parents and h) footwear item – parents). 0: Newly referred, 1: Clinical controls, 2: Pre-admission.

pre-admission children (group 2) would have a lower score in the physical domain than the clinical controls (group 1) was confirmed, based on the tendency from the box plots and a further unpaired *t*-test; the mean difference (95% CI) was 1.1 (0.4 to 1.8) (p = 0.002). The box plots indicated a tendency towards lower scores in the pre-admission children than in the other two groups in all three domains. In hypothesis testing, when assessing differences between groups it is important to assess whether any difference is as large as could be expected, rather than whether it is statistically significant, which depends on the sample size.[16]

The footwear item is included in the box plots for completeness, but it should be considered within the context of the individual child. The scores beyond the lower limits represent children with severe orthopaedic problems, who also had a low physical score, and should therefore not be considered outliers. All the stated hypotheses on convergent validity (Table VII) were proven in the child version. Although not moderate in the parent-proxy version, the correlation between the emotional domain and mental health was critical, and no critical correlation was proved between the emotional domain and self-esteem.

Feasibility. One item in one OxFQA completed by a child was incomplete (0.0008%). The mean time to complete the questionnaire was 3.0 minutes (0.75 to 7.0) for children and 2.1 minutes (0.25 to 5.0) for parents; 15 children (18.3%) needed their parents to read the OxAFQ out loud, and 67 (81.7%) completed it without assistance.

Examination of each domain revealed no floor effects, but ceiling effects were observed in the domains of school and play for 26 children (32%) and 23 parents (28%), and in emotional for 27 children (33%) and 19 parents (23%).

Table VII. Construct validity. Spearman's correlation coefficient for the Oxford ankle foot questionnaire-c with CHQ-CF87 (n = 65) and Oxford ankle foot questionnaire-p with CHQ-PF28 (n = 81) at the first test

Parents	Physical (p-value)	School and play (p-value)	Emotional (p-value)
Global health (n = 77)	0.33 (0.003)	0.24 (0.04)	0.14 (0.2)
Physical functioning (n = 78)	0.53* (< 0.001)	0.67* (< 0.001)	0.23 (0.04)
Role/social limitations – emotional/behavioural (n = 77)	0.28 (0.01)	0.24 (0.04)	0.17 (0.1)
Role/social limitations – physical (n = 78)	0.30 (0.007)	0.50* (< 0.001)	0.13 (0.3)
Bodily pain/discomfort (n = 78)	0.46* (< 0.001)	0.48* (< 0.001)	0.16 (0.2)
Behaviour (n = 78)	0.21 (0.06)	0.09 (0.4)	0.10 (0.4)
Global behaviour item (n = 78)	0.08 (0.5)	0.005 (1.0)	0.10 (0.4)
Mental health (n = 78)	0.13 (0.2)	0.15 (0.2)	0.34 (0.003)
Self-esteem (n = 78)	0.16 (0.2)	0.11 (0.4))	0.10 (0.4)
General health perceptions (n = 77)	0.06 (0.5)	0.01(0.9)	0.09 (0.4)
Change in health (n = 78)	0.23 (0.04)	0.13 (0.3)	0.16 (0.2)
Parental impact – emotional (n = 78)	0.36 (0.001)	0.30 (0.007)	0.28 (0.01)
Parental impact – time (n = 78)	0.18 (0.1)	0.25 (0.03)	0.11 (0.3)
Family activities (n = 78)	0.26 (0.02)	0.33 (0.003)	0.14(0.2)
Family cohesion (n = 75)	0.14 (0.2)	0.04 (0.7)	0.18 (0.1)
Children			
Global health (n = 61)	0.16 (0.2)	0.13 (0.3)	0.15 (0.3)
Physical functioning (n = 65)	0.65* (< 0.001)	0.67* (< 0.001)	0.28 (0.02)
Role/social limitations – behavioural (n = 65)	0.26 (0.03)	0.25 (0.05)	0.17 (0.2)
Role/social limitations – physical (n = 65)	0.43* (< 0.001)	0.40* (< 0.001)	0.29 (0.02)
Bodily pain/discomfort (n = 65)	0.67* (< 0.001)	0.60* (< 0.001)	0.10 (0.4)
Behaviour (n = 65)	0.30 (0.01)	0.22 (0.08)	0.42* (< 0.001)
Global behaviour item (n = 60)	0.28 (0.03)	0.19 (0.1)	0.16 (0.21)
Mental health (n = 65)	0.20 (0.1)	0.21 (0.09)	0.48* (< 0.001)
Self esteem (n = 65)	0.30 (0.01)	0.26 (0.04)	0.38 (0.002)
General health perceptions (n = 64)	0.17 (0.2)	0.10 (0.4)	0.29 (0.02)
Change in health (n = 65)	0.09 (0.5)	0.11 (0.4)	–0.03 (0.8)
Family activities (n = 65)	0.25 (0.04)	0.18 (0.2)	0.15 (0.2)
Family cohesion (n = 65)	0.17 (0.2)	0.21 (0.1)	0.20 (0.1)

* The moderate/strong correlations with significance

Discussion

Our study population resembles that of the developers in respect of mean domain scores, footwear item scores and heterogeneity.[1,2]

We found a high Crohnbach's alpha in both versions for all domains, although the emotional domain had a lower alpha of between 0.67 and 0.75, with the lowest coefficient in the OxAFQ-c. In general, we found a lower alpha than the developers,[2] and it was marginally lower in children than in parents, which is similar to the findings of the developers. The generally lower alpha and ICC among children in our study compared with that of the developers could be due to the method of administration of the questionnaires. In our study parents were explicitly instructed not to explain the questions to their children, whereas the developers permitted them to do so. Although this may occur in real life, we felt that it was a potential source of bias, overestimating reliability and agreement, as it might not reflect the children's own interpretation.

In keeping with former findings,[2] the physical domain scores were generally more discriminatory.

The retest questionnaire was completed in a different setting (at home), which might underestimate test–retest reliability. On the other hand, the heterogeneity of the diagnoses might overestimate reliability.

There are several strengths to our study compared with that of the developers. We had a larger sample size in the test–retest reliability assessment, we report a shorter range of response times, and the absolute errors of measurement were estimated for both versions of the OxAFQ, thereby elucidating the interpretation of the results when used in clinics.

Content validity in general is difficult to quantify. The experienced surgeons (OR, BMM) judged the properties of OxAFQ to be relevant and comprehensive for children with a variety of disorders of the foot and/or ankle.

Assessing construct validity was also challenging. We chose the CHQ as it was the only Danish PROM that had been thoroughly validated. Although validated in children with juvenile idiopathic arthritis, we believe it was a fair comparator for children with disorders of the foot and/or ankle. Hypothesis testing was limited by the health-related properties of CHQ, in contrast to the region-specific properties of the OxAFQ. This is a recognisable issue faced by other Danish studies validating knee-specific PROMs in children.[15] It was therefore unreasonable to expect more than a moderate correlation between similar domains in the two questionnaires. A potential risk of bias should be kept in mind when comparing Spearman's ρ in the hypothesis testing between children and parents owing to different

versions of the CHQ being used. Children in general found the emotional items in the CHQ demanding, and the questionnaire was tiring to complete.

A limitation to our study is that children aged < nine years (13 participants) were not asked to complete the CHQ, and therefore we did not validate the Danish OxAFQ in five- to eight-year-olds, only in their parents. Other paediatric and parent-proxy PROMs exist in Danish, but to our knowledge none have been validated. Despite this, we could have chosen an additional quick comparator such as Kidscreen[17] to validate the OxAFQ in children aged < nine years and perform a subgroup analysis, which could then be interpreted with caution. Based on clinical experience, we believe that in children aged between five and eight years the parent-proxy version would be a satisfactory and reliable method of assessing the impact of foot and/or ankle conditions. Morris et al[1] also suggested that although children aged < eight years can complete the questionnaire, their scores should be interpreted with caution.

The OxAFQ can discriminate between levels of severity of disease. The slightly negatively skewed scores among newly referred children might be justified by the observation that many of these did not have a problem that required orthopaedic intervention, which could also explain the observed ceiling effects in these domains. Although the differences in the domains of school and play and emotional are less obvious between the groups of children which we observed than those of the physical domain, the clinician is provided with relevant information on the child's perception of his/her wellbeing. Item 14 ('has anyone been unkind to you because of your foot or ankle?') tended to intercorrelate weakly with the other items in the emotional domain. It reflects the behaviour of others towards the child, and it could be argued that this item is culturally or socially dependent, thereby interfering with the developers' scaling of the emotional domain as one-dimensional. The domain is, thus, more of a formative rather than a reflective model, making Crohnbach's alpha difficult to interpret, as it is sensitive to the length of the domains and their dimensionality.[18] A short test and multi-dimensionality underestimates Crohnbach's alpha. Some researchers might advocate a factor analysis to assess the unidimensionality of the domains. We debated this subject with our statistician, who saw no reason for conducting factor analysis, as this too could give excessive empirical information not necessarily informative as to the internal consistency and unidimensionality of the test.

In conclusion, we found OxAFQ-c and OxAFQ-p to be valid in terms of agreement and reliability (internal consistency, test–retest reliability and measurement error), feasibility and construct validity. We also found OxAFQ-c and OxAFQ-p to be cross-culturally adaptable. The OxAFQ is easy and feasible to complete. To our knowledge, this is the first independent validation of the OxAFQ.

If it were to be used for evaluation purposes, we recommend that any trauma within the previous week be reported, as this could limit the validity of the answers. Future research is needed to determine the baseline values of healthy children, to reassess the items in the emotional domain and investigate the responsiveness of the OxAFQ.

Author contributions

P. Martinkevich: Design, Data collection, Data analysis, Writing the paper.
B. Møller-Madsen: Design, Data analysis, Writing the paper.
M. Gottliebsen: Design, Data analysis, Writing the paper.
L. Kjeldgaard Pedersen: Design, Data analysis, Writing the paper.
O. Rahbek: Design, Data analysis, Writing the paper.

We would like to thank G. Stougaard for her help with the administration of the questionnaires.

Children aged < nine years were not asked to complete the CHQ-CF87 as this was discussed in mail correspondence with Dr S. Nielsen, who performed the Danish validation of the CHQ, who suggested that this age group has difficulties understanding the questionnaire

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

This article was primary edited by E. Moulder and first proof edited by J. Scott

References

1. **Morris C, Doll HA, Wainwright A, Theologis T, Fitzpatrick R.** The Oxford ankle foot questionnaire for children: scaling, reliability and validity. *J Bone Joint Surg [Br]* 2008;90-B:1451–1456.
2. **Morris C, Doll H, Davies N, et al.** The Oxford Ankle Foot Questionnaire for children: responsiveness and longitudinal validity. *Qual Life Res* 2009;18:1367–1376.
3. **Hullmann, SE, Ryan JL, Ramsey RR et al.** Measures of General Pediatric Quality of Life. *Arthritis Care & Research* 2011;63:S420–S430.
4. **Nielsen S, Ruperto N, Herlin T, Pedersen FK; Paediatric Rheumatology International Trials Organisation.** The Danish version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). *Clin Exp Rheumatol* 2001;19(4Suppl23):S50–S54.
5. **No authors listed.** Isis Innovation: Isis outcomes translation and linguistic validation guidelines. Oxford: University of Oxford ISIS Innovation, in press 2015.
6. **Wild D, Eremenco S, Mear I, et al.** Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health* 2009;12:430–440.
7. **No authors listed.** John Martin Bland's Home Page, 2014. http://www-users.york.ac.uk/~mb55/index.html (date last accessed 01 December 2014).
8. **Bland JM, Altman DG.** Cronbach's alpha. *BMJ* 1997;314:572.
9. **Terwee CB, Bot SD, de Boer MR, et al.** Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
10. **McGraw KO, Wong SP.** Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30–46.
11. **Bonett DG.** Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–1335.
12. **Bland JM, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
13. **Mokkink LB, Terwee CB, Patrick DL, et al.** The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–745.
14. **Dawson J, Doll H, Coffey J, Jenkinson C; Oxford and Birmingham Foot and Ankle Clinical Research Group.** Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cartilage* 2007;15:918–931.
15. **Kocher MS, Smith JT, Iversen MD, et al.** Reliability, validity, and responsiveness of a modified International Knee Documentation Committee Subjective Knee Form (Pedi-IKDC) in children with knee disorders. *Am J Sports Med* 2011;39:933–939.
16. **Mokkink LB, Terwee CB, Patrick DL, et al.** The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–549.
17. **No authors listed.** The KIDScreen-10 index. http://www.kidscreen.org/english/questionnaires/kidscreen-10-index/ (date last accessed 30 January 2015).
18. **Tavakol M, Dennick R.** Making sense of Cronbach's alpha. *Int J Med* Educ 2011;2:53–55.

Paper II

BJR

RESEARCH

Precise and feasible measurements of lateral calcaneal lengthening osteotomies by radiostereometric analysis in cadaver feet

P. Martinkevich,
O. Rahbek,
B. Møller-Madsen,
K. Søballe,
M. Stilling

From Aarhus University Hospital, Denmark

Objectives

Lengthening osteotomies of the calcaneus in children are in general grafted with bone from the iliac crest. Artificial bone grafts have been introduced, however, their structural and clinical durability has not been documented. Radiostereometric analysis (RSA) is a very accurate and precise method for measurements of rigid body movements including the evaluation of joint implant and fracture stability, however, RSA has not previously been used in clinical studies of calcaneal osteotomies. We assessed the precision of RSA as a measurement tool in a lateral calcaneal lengthening osteotomy (LCLO).

Methods

LCLO was performed in six fixed adult cadaver feet. Tantalum markers were inserted on each side of the osteotomy and in the cuboideum. Lengthening was done with a plexiglas wedge. A total of 24 radiological double examinations were obtained. Two feet were excluded due to loose and poorly dispersed markers. Precision was assessed as systematic bias and 95% repeatability limits.

Results

Systematic bias was generally below 0.10 mm for translations. Precision of migration measurements was below 0.2 mm for translations in the osteotomy.

Conclusion

RSA is a precise tool for the evaluation of stability in LCLO.

Cite this article: *Bone Joint Res* 2015;4:78–83.

Keywords: Flatfoot; Calcaneal osteotomy; Radiostereometric analysis; Precision

■ P. Martinkevich, MD, PhD Student
■ O. Rahbek, MD, PhD, Associate Professor
■ B. Møller-Madsen, MD, DMSc, Professor
Department of Children's Orthopaedics, Aarhus University Hospital, Nørrebrogade 44, Building 1A, 2nd floor, 8000, Denmark.

■ K. Søballe, MD, DMSc, Professor, Department of Orthopaedic Surgery
Aarhus University Hospital, Tage-Hansens Gade 4, 8000 Aarhus C, Denmark.

■ M. Stilling, MD, PhD, Associate Professor, Department of Clinical Medicine
Aarhus University Hospital, Palle Juul-Jensens Boulevard 82, 8200 Aarhus N, Denmark.

Correspondence should be sent to Dr P. Martinkevich; email: polina.martinkevich@gmail.com

doi:10.1302/2046-3758.45.2000292 $2.00

Bone Joint Res 2015;4: 78–83
Received 16 September 2014; Accepted after revision 10 March 2015

Article focus

- Assessment of the precision of marker-based radiostereometric analysis (RSA) in lateral calcaneal lengthening osteotomies, with focus on the osteotomy site and the neighbouring calcaneal–cuboid (CC) joint.
- Double examinations were taken with different lengthenings of the osteotomy to mimic clinical situations.
- The feasibility of tantalum bead insertion in the tarsal bones was investigated.

Key messages

- The precision for migration measurements was good for translations within the ostetomy site and the CC joint.
- RSA is feasible within tarsal bones such as the calcaneus and the cuboid.
- This is the first study investigating the precision and feasibility of RSA in calcaneal osteotomies.

Strengths and limitations

- The study was conducted on fixed adult cadaver feet
- Three markers were visible in the cuboid and the anterior fragment of the calcaneus. Insertion of at least four markers would have been better
- Only the repeatability and feasibility were assessed, not accuracy. The settings were consistant and controlled.

Introduction

Radiostereometric analysis (RSA) is a highly accurate and precise method for quantifiying the three-dimensional movement between

rigid bodies, such as the stability of fracture parts and the migration of a prosthetic implant in the host bone.[1] It has gained great importance in the assessment of joint replacements and new operative techniques. RSA requires insertion of tantalum beads in the bony structures of interest in order to define distinct reference landmarks. The object (i.e. fracture or implant) and two layers of metallic markers in a calibration cage below the patient, are recorded in two simultaneous radiographs (stereoradiographs). During computer analysis of the stereoradiographs, the position of the object with respect to its reference landmarks is described as translations along, and rotations about, all three orthogonal axes in the Cartesian coordinate system. The lateral calcaneal lengthening osteotomy (LCLO) is a common operative approach for the correction of pes planovalgus in paediatric orthopaedic surgery. The lengthening is performed by insertion of a trapezoidal bone graft. For many years, autologous bone harvested from the iliac crest has been the preferred technique, and remains so in Denmark. Donor site morbidity, such as chronic pain, superficial infection, haematoma and lesions to the lateral femoral cutaneus nerve, has led to an increase in new structural substitutes to a bone graft.[2] Although there are numerous different substitutes available, the clinical evidence of the mechanical stability of these new materials is sparse. The safety and long-term outcome in orthopaedic surgery is yet to be assessed. A common radiological finding related to the LCLO is a 'subluxation' of the calcaneal–cuboid (CC) joint. The actual consequences of this finding are yet to be documented. Therefore, before further clinical studies with marker-based RSA on the tarsal bones are undertaken, we would need also to know the precision and feasibility of RSA on the CC joint.

When introducing new implants in hip and knee arthroplasty, the trend points towards using RSA as a quality tool in the phased clinical introduction of new prostheses to assess safety and provide better patient care.[3] To our knowledge, RSA has not yet been used in measurements of stability in foot surgery.

In preparation for a clinical trial to investigate the structural durability of an artificial bone graft *versus* an autologous bone graft in LCLO on paediatric patients and the CC joint, we conducted a RSA study on LCLO on cadaver feet.

The primary aim of this study was to investigate the precision and feasibility of the marker-based RSA system in a LCLO with different lengthenings, focusing firstly on the osteotomy site and secondly on the CC joint.

Materials and Methods

The main focus of this study was to assess the precision of RSA in calcaneal lengthening osteotomies. During the study preparation, the Standards for Reporting of Diagnostic Accuracy (STARD)[4] initiatives were followed.

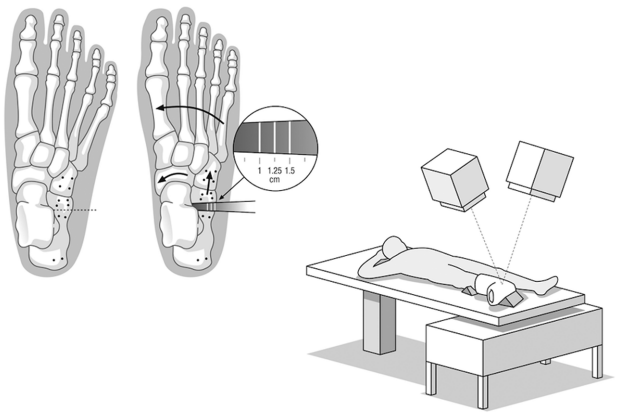


Fig. 1

Drawing showing the principles in a lateral calcaneal lengthening osteotomy with inserted tantalum markers (left) and the set-up of the radiostereometric analysis (right).

Cadavers. A total of three fixed cadaveric bodies were used, giving a total of six feet. They were all men of Caucasian origin. The mean age was 77 years (73 to 88), and all feet measured European size 40. There was no history of foot surgery. Causes of death were prostate cancer in two and unknown in the third. The cadavers were fixed according to a standardised procedure. The fixing fluid composition was made up of 10.5 L per cadaver of 5 L alcohol 96%, 2.5 L glycerine, 2 L destillated water and 1 L formaline. The cadavers were stored at 6°C, and the study was performed within one day.

Operative procedure and bead marking. The osteotomies were performed in an air conditioned room with temperature set to 22°C. The cadavers were allowed to temperate a minimum of 60 minutes prior to the operative intervention.

The LCLO was performed, following the exact procedure each time, by one experienced surgeon (OR) and one assisting surgeon (PM). A plexiglas trapezoidal wedge was specially manufactured for the study, and was marked corresponding to a lateral border lengthening of 1 cm, 1.25 cm and 1.5 cm (Fig. 1).

A lateral longitudinal incision was made over the calcaneus. Due to the stiffness of the tissue, blunt dissection was not performed, and instead a sharp incision was used straight to the bone. The capsule in the CC joint was respected. The periosteum was incised in line with the planned osteotomy, starting laterally approximately 1 cm to 1.5 cm proximal to the CC joint. The osteotomy was performed with an oscillating saw. Two osteotomes were used to distract and mobilise the osteotomy.

Following the osteotomy, in preparation for the RSA, spherical tantalum beads were inserted in the anterior and posterior calcaneal fragment and in the cuboid bone with a bead injector (Wennbergs Finmek, Sweden). The posterior calcaneal fragment was marked with four to six

tantalum beads at 1.0 mm each, the anterior calcaneal fragment was marked with three to four beads at 0.8 mm each, and the cuboid bone was marked with three to four beads at 1.0 mm each. Markers in the anterior and posterior calcaneal fragments were inserted through the osteotomy. Markers in the cuboid bone were inserted via a distal expansion of the initial skin incision through a small hole drilled in the lateral cortex with a 1.8 mm Kirschner wire. The position of the marker was controlled by intra-operative fluoroscopic x-rays. Finally, the cadavers were cautiously transported to the RSA examination room, where the final and gradual lengthening of the osteotomies by use of the plexiglas wedge and subsequent RSA imaging was performed.

RSA set-up. A standard set-up of two synchronised ceiling fixed x-ray tubes (Arco-Ceil/Medira; Santax Medico, Aarhus, Denmark) angled towards each other at 40º and a focused-grid uniplanar carbon calibration box (Box 24, Medis Specials, Leiden, The Netherlands) was used. The exposure was set to 70 kV, 10 mAs, fine focus. All radiographs were fully digital (FCR Profect CS, Fujifilm, Tvedbæk, Denmark) and they were stored without compression. The calibration box defined the position and orientation of the global coordinate system. The cadaver foot was parallel to the calibration cage, with the lateral side of the foot facing the calibration cage, the medial side facing upwards and the ankle in a neutral position. The focus was centred on the CC joint. All stereoradiographs were taken by a team of two radiographers.

Double-examination stereoradiographs were obtained with the calcaneal osteotomy in zero distraction, approximately 1 cm (wedge a), 1.25 cm (wedge b) and 1.5 cm distraction (wedge c). The foot was repositioned between each double examination. A total of 24 double examination stereoradiographs were obtained, four double examinations of each foot. Marker-based analysis of all stereoradiographs was performed by one observer (PM) with Model-based RSA 3.32 software (RSA*core*, Leiden, The Netherlands). The inclusion quality of the radiographs were condition numbers (spatial dispersion of markers within the bone fragment) for the marker models in the calcaneal fragments below 300 for each marker model, and a rigid body error threshold (marker fixation threshold) of maximum 0.5 mm. The condition number was obtained from experience with clinical marker dispersion in a similar sized bone in the human wrist.[5]

The higher the condition number, the poorer the marker dispersion, although condition numbers up to 758 have been reported for small anatomical locations, such as a model of a distal radius fracture.[6] One foot was excluded from the study, as the RSA data clearly indicated loose tantalum markers (high rigid body errors). A second foot was excluded from the precision analysis of the CC joint, due to a high condition number in the cuboid bone (up to 484).

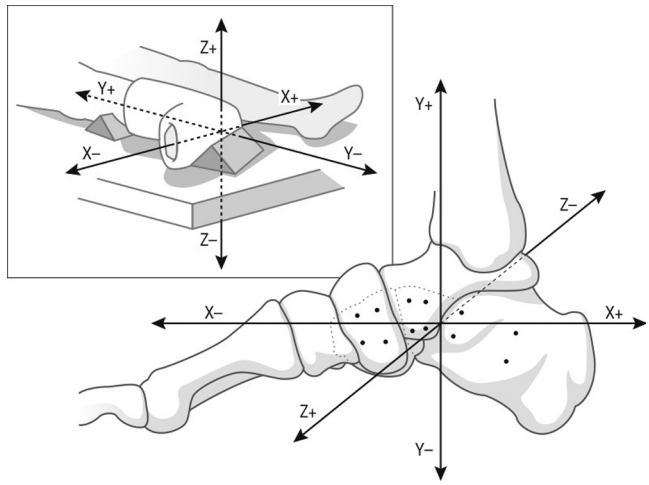


Fig. 2

Drawing illustrating the orientation of the coordinate system relative to the anatomy of the foot. The foot is viewed from the medial side, with the dotted lines representing the contour of the distal part of both the calcaneal bone and the cuboid bone. The drawing in the upper left corner illustrates the positioning of the foot for future clinical studies.

When assessing the precision of the osteotomy measurements, the centre of gravity of the marker model of the posterior calcaneal fragment made up the fixed rigid body reference and the marker model of the anterior calcaneal fragment made up the migrating object. For the CC joint, the centre of gravity in the anterior calcaneal fragment made up the fixed rigid body reference, and the marker model of the cuboid bone made up the migrating object. The marker models in the coordinate systems were oriented with the osteotomy performed in the y-axis, the lengthening in the x-axis, and any medial lateral movement of the migration object with reference to the rigid object would occur in the z-axis. A distraction of the osteotomy would be indicated by a negative sign in the x-translations and, conversely, a positive sign in the x-translations would indicate compression of the osteotomy (the anterior calcaneal fragment moving towards the posterior fragment) (Fig. 2).

Statistics and precision of RSA. The primary aim of this study was to discover the repeatability of zero migration measurements in the LCLO at different distraction points, as well as the further repeatability of zero migration measurements in the CC joint.

Repeatability/precision of the RSA set-up was assessed by 20 double examinations (16 in the CC joint) in neutral and with all three distractions of the osteotomy. The bias, which is the systematic variation of the double measurements, was calculated as the mean difference ($mean_{diff}$) between the first and second set of double measurements. The random variation was expressed as standard deviation of the differences ($SD_{diff\text{-}inter}$) between the first and second measurements (double examinations) along with 95% limits of agreement (LOA) according to

Table I. Repeatability of n = 20 double measurements in the calcaneal osteotomy for the anterior calcaneal fragment with reference to the posterior calcaneal fragment; combined precision for all three distraction grades

	Translation (mm)			Rotation (°)		
	x-axis	**y-axis**	**z-axis**	**x-axis**	**y-axis**	**z-axis**
Mean*	-3.33	-3.03	3.08	-5.34	6.41	3.04
$Mean_{diff}$†	-0.02	0.09	-0.02	0.07	0.19	-0.2
95% CI‡	(-0.05 to 0.01)	(0.04 to 0.15)	(-0.07 to 0.03)	(0.01 to 0.13)	(0.05 to 0.33)	(-0.46 to 0.05)
SD_{diff_inter}§	0.06	0.10	0.09	0.10	0.25	0.46
± LOA¶	0.11	0.20	0.17	0.21	0.50	0.90
p-value**	0.16	0.002	0.44	0.02	0.01	0.11

*Mean of the lengthening of the osteotomy
†$Mean_{diff}$ between the first and second set of double measurements, bias (systematic variation)
‡ 95% confidence intervals (CI) for the mean difference
§ standard deviation $(SD)_{diff\_inter}$, repeatability SD, random variation
¶ ± limits of agreement (LOA), 95% LOA, expected clinical precision
**p-value, paired *t*-test

Table II. Repeatability of n = 16 double measurements in the calcaneal–cuboid joint with the cuboid bone as the migrating object with reference to the anterior calcaneal fragment; combined precision for all three distraction grades

	Translation (mm)			Rotation (°)		
	x-axis	**y-axis**	**z-axis**	**x-axis**	**y-axis**	**z-axis**
Mean*	0.36	0.98	-1.54	2.50	-3.49	-0.54
$Mean_{diff}$†	0.01	-0.10	0.03	-0.06	0.09	0.33
95% CI‡	(-0.05 to 0.06)	(-0.18 to -0.03)	(-0.11 to 0.16)	(-0.24 to 0.13)	(-0.46 to 0.64)	(0.07 to 0.58)
SD_{diff_inter}§	0.095	0.12	0.24	0.31	0.93	0.43
±LOA¶	0.19	0.24	0.46	0.61	1.82	0.85
p-value**	0.77	0.01	0.72	0.53	0.75	0.02

*Mean of the lengthening of the osteotomy
†$Mean_{diff}$ between the first and second set of double measurements, bias (systematic variation)
‡ 95% confidence intervals (CI) for the mean difference
§ standard deviation $(SD)_{diff\_inter}$, repeatability SD, random variation
¶ ± limits of agreement (LOA), 95% LOA, expected clinical precision
**p-value, paired *t*-test

Bland and Altman.[7] The systematic variation should optimally be zero, and it reflects the precision in the cadaver RSA set-up, where LOA ($mean_{diff}$ ×$SD_{diff\text{-}inter}$ ±1.96) represents the expected clinical RSA precision for LCLO migration measurements.

Differences within and between double examinations (measurement one and measurement two) were tested for normality by QQ-plots. The mean differences within double examinations were tested for equal variance with an F-test, and if this assumption was fulfilled, a paired *t*-test was performed. The mean difference for each distraction grade (a,b,c) between groups was tested with an F-test for equal variance. If no significant differences were revealed, the mean differences were pooled.

Statistical computations were performed in Excel version 14.3.5 (Microsoft Excel for Mac 2011, Redmond, Washington) and Stata version 12.1 for Mac 2012 (StataCorp, College Station, Texas). The level of statistical significance was set at $p < 0.05$.

Results

The mean condition number was 53 (SD 32; 33 to114) in the posterior calcaneal fragment, 153 (SD 73; 82 to 268) in the anterior calcaneal fragment and 120 (SD 48; 88 to 199) in the cuboid bone. One foot had high condition numbers in the cuboid bone ranging from 471 to 484 and was, thus, excluded from statistical analysis.

We found significant differences in the means between some double examinations, but as they were very small and approximated zero, this was of no clinical relevance. Therefore, we pooled the data for all cadavers to get a common and more useful precision estimate.

The precision of the RSA measurements in the calcaneal osteotomy and in the CC joint is referred to in Tables I and II. The systematic variation in measurements of translations was below 0.09 mm in the osteotomy and below -0.10 mm in the CC joint. As for rotations, it was below 0.19° in the osteotomy and below 0.33° in the CC.

Precision was good for the calcaneal osteotomy measurements with LOA below 0.2 mm for translations (x,y,z) and below 0.9° for rotations (Fig. 3). Precision of measurements in the CC joint was below 0.46 mm for translations and below 1.8° for rotations (Fig. 4). As for rotations, the best precision was observed on the x-axis, with 0.21° for the calcaneal osteotomy and 0.61° for the CC joint. The poorest precision was observed on the z-axis for the osteotomy, at 0.9° and on the y-axis, 1.82° in the CC joint. Having mentioned this, the rotations are not of primary clinical interest.

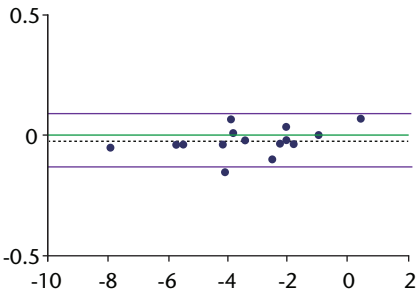


Fig. 3a

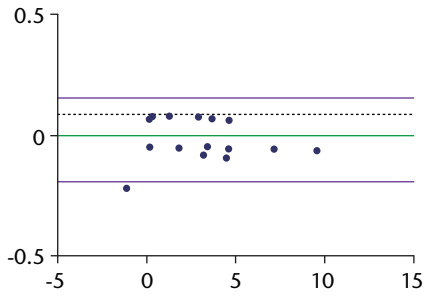


Fig. 3b

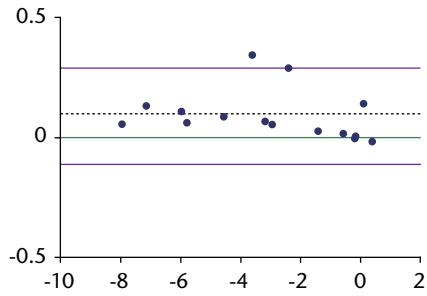


Fig. 3c

Bland-Altman plots of repeatability (precision) with 95% limits of agreement (LOA) for translations in the osteotomy for the x (a), y (b) and z (c) translations. The x-axis shows the average of two measurements (a double-examination) and the y-axis shows the difference between them. The green line denotes the optimal bias, which is zero. The dashed black line denotes the systematic bias from zero. The purple lines denote the 95% LOA. The blue dots represents measurements.

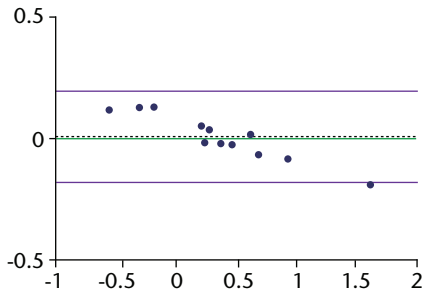


Fig. 4a

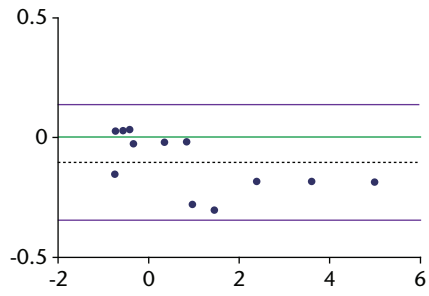


Fig. 4b

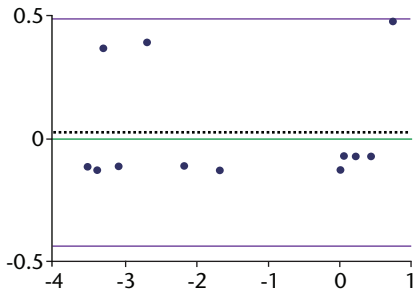


Fig. 4c

Bland-Altman plots of repeatability (precision) with 95% limits of agreement (LOA) for translations in the CC joint for the x (a), y (b) and z (c) translations. The x-axis shows the average of two measurements (a double-examination) and the y-axis shows the difference between them. The green line denotes the optimal bias which is zero. The dashed black line denotes the systematic bias from zero. The purple lines denote the 95% LOA. The blue dots represent the measurements.

Discussion

To our knowledge, these are the first published data on the precision and feasibility of RSA for evaluation of calcaneal lengthening osteotomies. Our main findings suggest a good precision for RSA measurements of translations in the calcaneal osteotomy and the CC joint with calcaneal lengthenings – as measured on the lateral side of the osteotomy – from 0 cm to approximately 1.5 cm. Rotation error was high on the z-axis in the CC joint, which would be expected as we are dealing with small bones and marker models and, thus, almost in-line bead insertion in the anterior calcaneal fragment. Error in rotation matrix increased with the increasing condition number,[8] which corresponds with findings in former studies examining small anatomical regions.[5,9] Therefore, it is crucial that the markers are as widely distributed as possible in a clinical study. In our planned clinical study, we aim to detect any translations in the CC joint in the y-axis after a LCLO, as this is a relatively common observation, and the clinical significance remains uncertain.

As a result of our experience in the present study, we provide the reader with some suggestions for future experimental and clinical studies using RSA on calcaneal osteotomies. When small neighbouring bones, such as those in the carpus or tarsus, are studied, we recommend marking neighbouring bones with different sized markers, as it enables the analyst to distinguish the location of markers in the oblique angled stereoradiographs, which in turn makes analysis easier and safer. During insertion of the beads, the anterior and posterior calcaneal fragments can be accessed easily through the osteotomy, thus avoiding further dissection of soft tissues close to the calcaneus. Insertion of beads in the cuboid and the most posterior calcaneal fragment requires either a larger skin incision than that needed for the calcaneal osteotomy, or a percutaneous bead insertion via fluoroscopic guidance. In the clinical setting, we would prefer a longer incision over additional skin penetrations with instruments that need to penetrate the bone, in order to avoid unnecessary risk of infection and damage to nerves, vessels and tendons.

In general, the surgical procedure and bead injection was challenging, but feasible, on fixed adult cadaver feet. We expect the RSA set-up to be easier to perform *in vivo*, and even easier on paediatric patients, who, despite their smaller bones, have elastic soft tissues and greater joint mobility, which eases access to the bones.

An important issue, which must be considered prior to further implementation of RSA in assessment of the stability of the calcaneal osteotomy performed in paediatric patients, is the dose of radiation to be used.

When one considers procedural variations within the planned clinical trial, the upper limit of the dose of radiation within our RSA system is no higher than 0.01 μSv. In comparison a person on a flight receives 0.005 mSv per hour in the air.[10] Based on these data, we conclude that RSA would be safe to use as a quality tool in calcaneal lengthening osteotomies in paediatric patients.

Limitations

The study was designed to test the precision of RSA for measurements of stability in LCLO in preparation for a clinical study on paediatric patients. Ideally, the study should have been performed on paediatric cadaver feet. The feet used for this study were normally configured, except for one pes equinus foot. Adult feet are different in bony structure from paediatric feet, with a bigger anatomy and more rigid joints and soft tissues. The cadaver feet were fixed, which demanded a greater effort to access and distract the osteotomy site due to the stiff tissues and muscles. The distraction was more difficult to perform in cadaver feet than it would be on vital human paediatric feet, as we had to be careful not to fracture the fragile cadaver bone. Therefore, lengthening of the osteotomy according to the markings on the plexiglas wedge was not similar for all cadaver feet. We emphasise that this study was not intended as an accuracy study, due to the obvious differences between adult cadaver feet and paediatric feet. We would expect any measured accuracy to be higher in clinics, and it would be an interesting issue to assess.

Bony structures should be marked with at least three non-collinear markers,[5] but four was considered to be much better, yielding more markers for data processing in case of problems such as marker loosening, linear marker distribution or marker occlusion. We used three to four markers in the cuboid bone and the anterior calcaneal fragment, and it was difficult to insert more markers without risking placement of very close markers or touching/loosening markers that were already placed. Therefore, in the clinical setting, we do not think it will be possible to place more than four markers in these small bones and bone fragments safely.

In conclusion, we found RSA to be a feasible method, with good precision for measurements of stability in the osteotomy and of any movement in the CC joint. Measurements of rotation were less precise, which is to be expected in small bones.

Acknowledgement: We thank the Institution of Biomedicine - Anatomy, Aarhus University for their contribution with the cadaver feet and the Department of Pathology at Aarhus University Hospital, Tage-Hansens Gade, for use of facilities for preparation and safe-keeping of the cadavers. Also, we thank the Department of Radiology at Aarhus University Hospital, Tage-Hansens Gade, for their experienced assistance with the radiographic examinations.

Supplementary material

Tables showing significant differences in means between double examinations are available alongside the online version of this article at www.bjr.boneandjoint.org.uk

References

1. **Valstar ER, Gill R, Ryd L, et al.** Guidelines for standardization of radiostereometry (RSA) of implants. *Acta Orthop* 2005;76:563–572.
2. **Kurien T, Pearson RG, Scammell BE.** Bone graft substitutes currently available in orthopaedic practice: the evidence for their use. *Bone Joint J* 2013;95-B:583–597.
3. **Nelissen GHH, Pijls BG, Kärrholm J, et al.** RSA and Registries: The quest for phased introduction of new implants. *J Bone Joint Surg [Am]* 2011;93-A(Suppl 3):62–65.
4. **Bossuyt PM, Reitsma JB, Bruns DE, et al.** The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. The Standards for Reporting of Diagnostic Accuracy Group. *Croat Med J* 2003;44:639–650.
5. **Hansen TB, Stilling M.** Equally good fixation of cemented and uncemented cups in total trapeziometacarpal joint prostheses: a randomized clinical RSA study with 2-year follow-up. *Acta Orthop* 2013;84:98–105.
6. **Madanat R, Mäkinen TJ, Moritz N, Mattila KT, Aro HT.** Accuracy and precision of radiostereometric analysis in the measurement of three-dimensional micromotion in a fracture model of the distal radius. *J Orthop Res* 2005;23:481–488.
7. **Bland JM, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
8. **Söderkvist I, Wedin PA.** Determining the movements of the skeleton using well-configured markers. *J Biomech* 1993;26:1473–1477.
9. **Ryd L, Yuan X, Löfgren H.** Methods for determining the accuracy of radiostereometric analysis (RSA). *Acta Orthop Scand* 2000;71:403–408.
10. **No authors listed.** American Nuclear Society. http://www.ans.org/pi/resources/dosechart/msv.php (date last accessed 24 March 2015).

Funding statement:
- None declared

Author contributions:
- P. Martinkevich: Study design, Surgery, Data collection, Data analysis, Writing the draft paper, Revising the paper
- O. Rahbek: Study design, Surgery, Revising the paper
- B. Møller-Madsen: Study design, Surgery, Revising the paper
- K. Søballe: Study design, Revising the paper
- M. Stilling: Study design, Data collection, Data analysis, Revising the paper

ICMJE Conflict of Interest:
- None declared

Paper III

Structural HATCP graft versus tricortical iliac crest autograft in paediatric calcaneal lengthening osteotomies.

Interim results from a randomised, controlled non-inferiority study

P. Martinkevich[1], O. Rahbek1, M. Stilling[2], L.K. Pedersen[1], M. Gottliebsen[1], K.Søballe[3], B. Møller-Madsen[1]

[1]Dept. of Children's Orthopaedics, Aarhus University Hospital, Denmark

[2]Dept. of Clinical Medicine, Aarhus University, Denmark

[3]Dept. of Orthopedic Surgery, Aarhu University Hospital, Denmark

Abstract

Aims: To compare the structural durability of hydroxyapatite-tricalcium phosphate (HATCP) to autologous (AUTO) iliac crest bone graft in calcaneal lengthening osteotomies (CLO) for paediatric pes planovalgus.
Patients and Methods: We present interim results of 11 patients (HATCP, n=6 and AUTO, n=5) with a mean age of 11.5 years (range 8.2-14.2) from a randomised controlled non-inferiority trial with 6 months follow-up. The primary outcome was the stability of the osteotomy measured by radiostereometric analysis. A non-inferiority margin of ≤2 mm osteotomy compression was set.
Results: Six months follow-up data revealed a compression in the osteotomy by a mean 1.97 mm (99.8%CI: -1.65; 5.60) greater in the HATCP-group compared to the AUTO-group. Migration in the CLO grafted with HATCP stabilised at six months follow-up as compared to six weeks follow-up with AUTO-graft.
Conclusion: This is the first randomised trial investigating the efficacy of HATCP-graft in comparison with AUTO-graft concerning stability of CLO in paediatric patients. The drawbacks outweighted the strengths of the HATCP and thus, further inclusion of patients was terminated. We do not recommend HATCP-graft for use in unfixed CLOs in its current structure.

Introduction

The calcaneal lengthening osteotomy (CLO) is a common surgical technique for correction of flexible pes planovalgus (PPV) in paediatric patients.[1] Lengthening is achieved by insertion of a structural trapezoidal shaped tricortical iliac crest bone graft. Typically, a structural autograft or allograft[2] is used. Among the major drawbacks related to using autograft is the inherent donor site morbidity, especially donor site pain.[3,4] Other limitations to using allografts in children include risk of infection,[5] limited availability, high costs and ethical concerns with use of donor bone in children.

Calcium phosphate bioceramics (CPBCs) are promising alternatives to autograft owing to their biological safety, absence of foreign body reaction, their osteoconductive capacity and ability to facilitate osseointegration.[5,6] Their porous and brittle nature has been the major limitation in their mechanical load bearing potential. However, the latest bone tissue engineering research has focused on improving the mechanical features of CPBCs [7] in pursuit of an optimal and tailored composition of tricalcium phoshpate (TCP) and hydroxyapatite (HA).

We considered a new indication for HATCP as a potential bone graft substitute for CLO in paediatric patients for three major reasons: 1) The calcaneal bone is lengthened by no more than 1-1.5 cm on the lateral side. 2) The foot is immobilised for a minimum of six weeks with no weight-bearing, and we expect this to be sufficient time to allow for native bone ingrowth. 3) Children and adolescents are considered to have an efficient bone healing potential.[8]

To date there is no clinical randomised trial comparing structural HATCP with tricortical iliac crest autograft (AUTO) in CLO in paediatric patients with flexible PPV. Previous studies comparing structural bone grafts in paediatric foot surgery[9,10] have used bone healing assessed by conventional X-rays, which is subject to observer variability.[11] Furthermore, different criteria are used to define the rigid fixation/structural durability of bone graft,[9,10,12,13] which makes it difficult to generalise the results.

Radiostereometric analysis (RSA) is an accurate and precise method for continuous three-dimensional measurements, and has shown feasible and precise assessment of migration between the distal (dCF) and proximal calcaneal fragments (pCF) in CLOs.[14]

We hypothesised the HATCP-graft would be non-inferior to AUTO-graft in maintaining the lengthening of CLO in paediatric patients with flexible PPV by ≤2 mm osteotomy compression measured as x-translations by RSA. In this paper we present the interim results of the non-inferiority trial.

Methods and Patients

Ethics

The study was registered within the Data Protection Agency (J. No. 1-16-02-86-12) and ethical approval was obtained form the Central Denmark Region Committees on Biomedical Research Ethics (Record No. 1-10-72-250-12, Appendix No. 40345). The protocol was registered in ClinicalTrials.gov (NCT01770574) and the study was conducted in accordance with the declaration of Helsinki 1995.

Design

Interim results from a randomised controlled two-group parallel non-inferiority design with equal randomisation ratio of 1:1, conducted and reported in accordance with the Consort Guidelines for Reporting of Non-inferiority trials. [15]

Patients

Patients with symptomatic flexible PPV scheduled for CLO at our institution were considered eligible for inclusion if they met the following criteria: Age 5 to 16 years at the preassessment date, no major cognitive impairments, intact ambulatory function, no major simultaneous ipsilateral orthopaedic interventions. Patients would be excluded in case of reoperation.

Enrollment took place at the Paediatric Orthopaedic Outpatient clinic from 2012 to 2015 by the primary investigator (PM). After obtaining written consent patients were randomised to receive either HATCP or AUTO-graft. Concealed sequential allocation based on surgery date was used with opaque envelopes numbered 1-30. Patients were allocated to the intervention by the primary investigator after onset of anaesthesia and were thus not disclosed to the intervention until after the operation.

Bone graft substitute

The HATCP-graft provided by Ceramisys Limited (Sheffield, England) is a biphasic calcium-ceramic with a 60/40 ratio composition of hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$, HA and β-tricalciumphosphate (β-$Ca_3(PO4)_2$),

β-TCP. It is a trapezoidal shaped block, manufactured in different sizes. HATCP has fully interconnected pores, with a porosity of 83 %, macroporosity within 250-800 μm and a microporosity of 1.0-10 μm. The sintering temperature is above 1200 °C and the compressive strength is 4-7 MPa (information provided by Ceramisys Limited).

Operative procedure

The operation (Appendix 1) followed a modified Mosca procedure.[16] A longitudinal skin incision was made just proximally to the calcaneo-cuboid joint (CCJ) (Figure 1). The incision was expanded further distally over the cuboid to gain access with the bead injector for the subsequent bone markings. The ostetomy was performed about 1.5 cm proximally and parallel to the CCJ with an oscillating saw. By use of a laminar spreader the osteotomy was dilated until correction of the deformity was achieved (10-12 mm). If the patient was allocated to autograft, iliac crest bone graft harvesting was approached by a minimum 5 cm anterolateral incision 1 cm below the iliac crest, performed simultaneously by an assisting surgeon.

If the HATCP bone graft substitue was used it was often necessary to trim the graft after insertion as the graft tended to protrude laterally. This produced debris, which was distributed in the surrounding soft tissue and difficult to remove.

In preparation for RSA the dCF and pCF were marked with at least four tantalum beads, which were as widely distrubuted as possible using bead injectors (Kulkanon, Wennbergs Finmek AB, Sweden) of matching calibers. The dCF and pCF were accessed through the osteotomy and from the lateral side. The dCF was marked with 0.8 mm beads and pCF were marked with 1.0 mm beads. The bead and graft position were controlled by fluoroscopy.

The foot was immobilised in a below knee X-lite circular cast, with the ankle in neutral.

No weight-bearing was allowed for six weeks. Subsequently, weight-bearing was allowed until pain tolerance.

Preoperative antibiotic administration adhered to a standard regime with dicloxacilline or cefuroxime single-dose intravenous injections. The standard analgesia administration was for the iliac crest a subcutaneous injection of 10 mL ropivacain (concentration 7.5 mg/mL) and for the osteotomy site a nerveblock (saphenus and tibial nerve) including a single-shot and a catheter with continuous infusion of body-weight adjusted ropivacain (2mg/mL), rate 3-5 mL/hour.

Figure 1.

Radiostereometric analysis set-up

A standard set-up of two synchronised ceiling fixed X-ray tubes (Arco-Ceil/Medira; Santax Medico, Aarhus, Denmark) angled towards each other at 40° and a focus-grid uniplanar carbon calibration box (Box 24, Medis Special, Leiden, The Netherlands) was used (Figure 2).

Figure 2.

Supine stereoradiographs were obtained within two days after surgery (baseline), at six weeks, eight weeks, six months and 12 months. Double-examinations were scheduled at six weeks follow-up, to assess the clinical precision/repeatability of the RSA measurements. Patients were repositioned between each double-examination.

Marker-based analysis of all stereoradiographs was done by one observer (PM) using Model-based RSA 3.32 software (RSA*core*, Leiden, The Netherlands). To ensure consistency in migration data, the analysis criteria requires the condition number (CN) and marker number to be the same within each individual marker-model throughout all scenes (X-ray pairs) and a rigid body error threshold of maximum 0.5 mm. The upper limit for CN was set to 350.[14]

Figure 3.

Repeatability

The repeatability of the RSA system was assessed in accordance with the International Standard and definitions (ISO1998). The systematic variation (bias) of the RSA system should optimally be zero, and is represented by the difference of means between double-examinations. The standard deviation of the mean differences reflects the random variation of the RSA measurements. The detection limit, or coefficient of repeatability (CR; $\pm 1.96 \times SD_{diff}$) represents the lower limit of the actual migrations. [17,18]

Secondary outcomes

Supine lateral and oblique X-rays were obtained six weeks after surgery to assess signs of bone graft replacement (trabecular continuity, loss of radiopacity and gradual blurring of the sharp margins) prior to cast removal. The lateral talo-1st-metatarsal angle (TMA), lateral calcaneal pitch angle (CP) and lateral talo-calcaneal angle (TCA) were measured as proposed by Davids et al.[19] The arch index (AI) was calculated according to Cavanagah et al.[20] Clinical outcomes were evaluated by the Oxford Ankle Foot Questionnaire (OxAFQ),[21] standard clinical examination, complications, postoperative pain, analgesics consumption, length of surgery and admission.

Statistics, interim guidelines and stopping guidelines

Sample size calculations were based on the non-inferiority continuous outcome trial method Sealed Envelope Ltd. 2012.[22] With a significance level of 0.05 (one-sided 95% CI), a power of 0.9, standard deviation of 1.5 mm in both groups and non-inferiority limit (Δ_{NI}) of 2 mm x-translation, the required sample size was 10 patients in each group. The Δ_{NI} was based upon the senior surgeons estimate of how much more compression could be accepted of the HATCP compared with the AUTO-graft. Hence, the primary outcome was the mean change in x-translation (Tx) of the dCF with reference to the pCF, analysed by a per-protocol (PP) approach.

This paper is based on an interim analysis, as defined by reaching five patients in each arm with six months follow-up. The primary outcome was analysed with relation to the Δ_{NI}. Non-inferiority was claimed if the upper limit of a two-sided 99.8% CI would be below the Δ_{NI} of 2 mm (HATCP graft minus AUTO graft osteotomy migration).

In case the HATCP material would appear to be worse than the AUTO graft, predefined stopping guidelines were set: The primary outcome measure was the mean difference in x-translation in the osteotomy

(osteotomy migration) in the two groups (HATCP migration minus AUTO migration). The statistical support team advised to consider analysing the interim results from a 99.8% CI. If the lower limit would be above zero and the upper limit would be above the Δ_{NI} of 2 mm, or the entire 99.8% CI would lie above 2 mm, this would indicate inferiority of the HATCP to AUTO graft.[15] It would therefore be advisable to stop the study to prevent adverse effects. If any complications requiring surgical intervention (Grade III)[23] with clear relation to the HATCP-graft, would occur, this would also advocate study termination.

Sensitivity analysis: The robustness of the primary outcome measure results was cross-checked by computing both a PP and an intention-to-treat (ITT) analysis imputing the last available results of the primary outcome measure (and also Ty, Tz) for the interim follow-up. Two-sided 99.8% CI were compared with one-sided 99.9% CI.

Data were approached by parametrical tests if prerequisite assumptions were fulfilled. Gaussian data distribution was tested by Shapiro-Wilk's test and equal variance was checked by f-test. Unpaired t-tests were used for group comparisons. Statistics were computed in Stata/IC 12.1® for Mac (StataCorp, USA). Repeatability results are presented with 95% CI.

Results

Participant flow is presented in Figure 4. Baseline characteristics for participants are presented in Table 1.

Figure 4.

Table 1. Demographics, reported as mean and range (n=11).

Patient variables	**HATCP group**	**AUTO group**
Subjects (n)	6	5
Age (years)	12.3 (9.9-14.2)	10.8 (8.2-13.7)
Male/female	1/5	2/3
Height (cm)	156.6 (132-164)	144.4 (137-162)
Weight (kg)	49.4 (33-65)	35.18 (28-47.7)
Right/left	2/4	2/3
PETA	1	3
Arch index	0.3(0.1-0.4)	0.3(0.1-0.4)
Preop X-rays	**HATCP**(n=3)	**AUTO** (n=4)
[b]Lat Talo-1M (°)	14 (9-22)	10 (7-12)
[c]Lat CP(°)	15 (13-18)	14 (10-20)
[d]Lat TC(°)	42 (33-52)	43(29-50)
Other diagnoses: Idopathic juvenile psoriatic arthritis (IJPA, 1), mild psychomotor retardation (2), cerebral palsy (hemiplegia, 2) and concomitant contralateral eight plate removal and revision of iliac crest scar from former surgery (1).		
[a]Arch index = $CA_{midfoot}/(CA_{midfoot}+CA_{hindfoot}+CA_{forefoot})$[20]		
[b]Lateral talo-1st-metatarsal angle, normal 13°±8[19]		
[c]Lateral calcaneal pitch, normal 17°±6[19]		
[d]Talo-calcaneal angle, normal 49°±7[19]		

Complications

HATCP-group

Four out of six patients had wound reactions at cast removal. Of these, two patients later required scar revisions under general anaesthesia due to hypergranulation. Two patients (including the patient who

received scar revision surgery) received oral antibiotics for 10 days after cast-removal. One patient had redness around the surgical wound at six weeks and later underwent revision surgery because of deep infection with removal of graft with identified pathogens Staphylococcus aureus and Coagulase-negative staphylococci (CoNS).

AUTO-group

Three patients had a hypertrophical hip scar; the donor site pain and scar tissue had resolved within three months follow-up in two patients and persisted in one. Another patient had a suspected foot wound infection which resolved after a 10 day oral antibiotics treatment.

All patients reported the postoperative pain at the hip as worse than in the operated foot. There was no difference between duration of surgery or length of admission between the two groups (Table 2).

Table 2. Postoperative data reported for patients included in interim-analysis (n=11). Data listed as mean(range).		
Postoperative variables	**HATCP** (n=6)	**AUTO** (n=5)
Postoperative pain numeric scale (0-10)		
Baseline	0.7 (0-4)	0.8 (0-2)
24 hours pain, mean	1 (0-3)	2.8 (0-6)
24 hours pain, max	5.8 (0-9)	5.6 (3-8)
Morphine equivalents/kg body mass (mg, po)		
24 hours	0.5 (0.2-1)	0.6 (0.3-1)
Discharge (total)	1.8(0.4-4.4)	1.7(0.4-4)
Total days of admission	2.6(1-4)	2.8 (2-4)
Duration of surgical time (hours)	1.6(1.3-1.8)	1.7(1.4-2.2)
Complications	**HA/TCP** (n=6)	**AUTO** (n=5)
Grade I		
Minor wound issues	n=2 *(foot)*	n=3 *(donor site)*
Grade II		
Antibiotic treatment of wound infection (foot)	n=2	n=1
Grade IIIb		
Reoperation or wound revision	n=3	

The gradings are based on the new and reliable complication classification system by Dindo et al [23].

Interim analysis of RSA measurements

Precision was good for osteotomy translations (Table 3).

Table 3. Precision/repeatability of n=10 double RSA measurements of the osteotomy model.			
N=10	Tx	Ty	Tz
$Mean_{diff}$	0.00	-0.03	0.03
SD_{diff}	0.03	0.05	0.07
95%CI[a]	-0.02;0.03	-0.04;0.03	-0.02;0.08
CR[b]	0.07	0.09	0.13

[a]95%CI: 95% confidence interval for $Mean_{diff}$

[b]CR: coefficient of repeatability, $\pm 1.96 \times SD_{diff}$

From Figure 5a graft-compression (Tx) is more pronounced in the HATCP-group compared to the AUTO-group, and is beyond the detection limit for 5 out of 6 HATCP-cases. On the contrary, 4 out of 5 AUTO Tx are below the detection limit. The Ty are larger in the HATCP-group (Figure 5b). Tz migrations are insignificant (Figure 5c). The osteotomy stabilised at six months in the HATCP-group and at six weeks in the AUTO-group.

Figure 5a. x-translation

Figure 5b. y-translation

Figure 5c. z-translation

The osteotomy translations with mean and range are presented in Table 4.

Table 4. Mean(range) of Tx, Ty, Tz (in mm) in the osteotomy model within the two groups (n=10) at 6 months and n=8 (5 HATCP, 3 AUTO) at 12 months follow-up.

HATCP	6 weeks	8 weeks	6 months	12 months
Tx	1.07(-0.24;1.76)	1.44(-0.28;1.13)	2.22(-0.21;3.61)	2.12(-0.33;3.65)
Ty	1.87(0.62;3.63)	2.37(0.85;4.14)	4.41(1.50;7.77)	4.12(1.55;7.29)
Tz	-0.70(-1.77;0.28)	-0.73(-2.40;0.60)	-0.49(-1.96;0.99)	-0.45(-1.98;1.34)
TT[a]	2.48(0.66;4.24)	3.14(0.90;4.81)	5.25(1.59;8.04)	4.96(1.68;8.20)
AUTO	6 weeks	8 weeks	6 months	12 months
Tx	0.14(-0.46;1.2)	0.18(-0.45;1.43)	0.25(-0.44;1.93)	0.21(-0.5;1.34)
Ty	1.44(-0.03;2.06)	1.33(0.13;2.27)	1.52(0.27;2.70)	1.24(0.47;2.12)
Tz	-0.64(-1.6;0.14)	-0.60(-1.77;-0.008)	-0.61(-1.35;-0.18)	-0.17(0.4;0.12)
TT	1.47(0.14;2.87)	1.58(0.14;3.22)	1.81(0.46;3.58)	1.54(0.65;2.19)

The excluded patient had migration data until 8 weeks follow-up: Tx=5.04 mm, Ty=3.09 mm, Tz=-1.13 mm (6 weeks) and Tx=5.82 mm, Ty=5.32 mm, Tz=7.83 mm (8 weeks).

[a]Total translation = $\sqrt{(Tx^2+Ty^2+Tz^2)}$

The difference in graft compression between the two groups was 1.97 mm (99.8%CI: -1.65; 5.60, one-sided 99.9%CI: -2.08;6.03) and by including the reoperated patient the ITT analysis yielded a difference of 2.57 mm (99.8%CI: -1.65; 6.80). Total translations (TT) at each follow-up are presented for each group (Table 4). This reflects the total migration, defined by the 3D Pythagora's theorem, of the graft. The TT of HATCP was 3.45 mm (99.8%CI: -2.89;9.8). Although, not statistically significant, the confidence interval is skewed towards the right, supportive of a greater migragion of the HATCP-graft.

Each group had a patient with a contrasting migration pattern. In the HATCP-group, one migration pattern was below the detection limit for x-translation. This patient received NSAIDs and immunosuppresive medication due to psoriatic arthritis, but whether this is of relevance is not obvious. In the AUTO-group one patient had greater migration numbers. This patient had hemiplegia and the lateral column still appeared short after cast removal.

Secondary outcomes

Signs of bone graft replacement were present in all cases at six week follow-up assessed by conventional X-rays.

Both groups showed comparable changes towards higher OxAFQ-scores (improvement) in the physical function and footwear item at six months follow-up (Table 5).

Table 5. Mean(range) domain percentage of the OxAFQ scores in the HATCP and the AUTO group. There were 5 parent- and 3 children-reports in the HATCP group (two children were not able to respond to the questionnaire) until 12 months follow-up. Reports from the AUTO groups are available for 5 parent-children pairs until 6 months.

	HATCP			AUTO		
Domains	Baseline	6 months	12 months	Baseline	6 months	12 months
Physical						
Parents	32.6(25-54.2)	55.8(29.2-83.3)	70.8(45.8-87.5)	45.8(25-83.3)	79.2(37.5-95.8)	87.5(62.5-100)
Children	28.1(16.7-37.5)	55.6(50-58.3)	56.9(50-66.7)	63.3(50-70)	81.7(54.2-100)	93.1(91.7-95.8)
School & Play						
Parents	45.8(25-68.8)	68.8(56.3-87.5)	77.5(43.8-100)	72.5(56.3-100)	82.5(25-100)	91.7(75-100)
Children	40.6(31.3-43.8)	64.6(50-75)	58.3(43.8-87.5)	82.5(62.5-100)	88.8(62.5-100)	95.8(87.5-100)
Emotional						
Parents	67.7(43.8-87.5)	91.3(68.8-100)	88.8(75-100)	80(50-100)	83.8(43.8-100)	72.9(18.8-100)
Children	64.1(43.8-75)	91.7(81.3-100)	93.8(93.8-93.8)	90(75-100)	85(56.3-100)	79.2(43.8-100)
Footwear						
Parents	20(0-25)	60(0-100)	70(50-100)	40(0-75)	70(0-100)	66.7(0-100)
Children	25(0-50)	41.7(25-50)	50(25-75)	45(25-75)	70(25-100)	58.3(0-100)

Discussion

The primary outcome measure in this interim analysis was CLO stability at six months evaluated by radiostereometry as x-translation of the dCF with reference to the pCF, as an indirect measure of the structural durability of the bone graft materials. Available 12 months follow-up data were analysed. The HATCP-scaffold did not attain a stable fixation within the osteotomy at six weeks, as patients had further accentuation of osteotomy migrations after initiation of weight-bearing until six months as compared to the AUTO-group, where stability was evident at six weeks. The difference in compression between groups at six months was 1.97 mm (99.8%CI: -1.65; 5.60) in favour of AUTO, and robustness was confirmed by a sensitivity analysis.

Statistically, the results are non-significant and inconclusive regarding non-inferiority as the 99.8% CI includes zero and the noninferioriy-limit. The upper limit of the 99.8% CI exceeds the determined non-inferiority margin of the HATCP. The x-translations for the AUTO-group are below the detection limit in 4 out of 5 patients, thus we cannot conclude whether these are actual migrations.

The delayed fixation of HATCP is in agreement with a recent feasibility study on triphasic calcium ceramics (HASi) for various surgical procedures in a paediatric population. The authors found autograft to have a "incorporation time" of three months compared to six months for the HASi-graft (p=0.001). They reported malunion(1), non-union(4) and no local wound complications.[12]

Pain, admission length, surgical time and the difference in mean domain scores of the OxAFQ did not differ between the two groups in our study. Overall pain was scored, and subsequently the worst pain was identified (donor site or foot pain). All patients in the AUTO-group rated the donor site pain as the worst, but it did not yield a significant difference in mean pain scores between the two groups. The method by which pain was measured is liable to bias: There was no common standardised postoperative medication regime as a results of unexpected events such as non-functioning tibialis catheter, pressure from cast, pain from intravenous procedures and different sensitivity to analgesics. Additional factors, such as anxiety from e.g. being in an unfamiliar environment has also been found to influence the perception of pain.[24]

A recent prospective study investigated donor site morbidity following anterior iliac crest bone grafting in 33 children and adolescents (34 hips).[25] A Visual Analog Scale score was used to differentiate between the pain at the donor site and the recipient site, and the authors established a positive correlation between the measurements. Pain was more intense in patients who received flat foot surgery (autograft harvest) than other corrective osteotomies. We would expect that the overall pain would be more intense in patients who are exposed to several surgical interventions simultaneously.[26] Clarke et al[25] found no complications to graft harvesting, except for a one lateral cutanoeus femoral nerve injury which resolved spontaneously at three months. In our study only one patient had persistent hip pain at final follow-up because of hypertrophic scar tissue, and we suspect that the morbidity associated with iliac crest bone graft harvesting is overexposed in literature. Instead the perspective should be aimed towards optimising the pain management, especially in patients who are expected to be sensitive to postoperative pain.[27] Moreover, an effective treatment of hypertrophic scar tissue is yet to be established. [28]

The anticipated advantages of the HATCP-graft are limited by feasibility issues. The brittle nature of the material challenges preparation and insertion in the osteotomy. The graft debris easily dissipates into the surrounding soft-tissue making it difficult to wash out. Vertical cracks can occur in the HATCP-graft during the insertion, necessitating repetitive lengthening attempts. We suspect the wound inflammations leadning to antibiotic treatment and revisional surgery in the HATCP-group (Table 2) might be due to the dissipitation of the material into the surrounding tissue. Recent results from a histological investigation of tissue reactions after biphasic calcium phosphate in oral implamantology, have suggested that premature exposure of the material might carry a risk of inflammatory tissue reaction and impede new bone formation.[29] We are uncertain whether there is any association between the HATCP and the documented CoNS infection. CoNS are typical opportunists, and they are associated with implanted foreign bodies.[30]

The interim analysis should be cautiously interpreted, as data are incomplete, the sample size is small and the CI for treatment effect is wide. From an ethical stand point, should the new treatment be suspected to be worse, then weaker evidence is required to stop a trial[31].

From the stopping guidelines it would be adviseable to terminate the study if there was a clear association between the new bone graft material and the need for revisional surgery. We cannot provide statistical evidence for this association, however as one patient from the HATCP-group was reoperated, and two patients underwent in this group underwent surgerical revision under general anasthesia, and the interim results did not produce the stability results of the osteotomy in the HATCP group that would make it viable, it was difficult to justify continuing the study under the given conditions. Perhaps the HATCP would perform better if the osteotomy had been stabilised, as advocated by Mosca. This would require another study.

For future studies a data monitoring committee should be incorporated in the protocol.

In this study, a clinical decison was made based on available clinical and statistical data from the interim analysis to terminate the study.

It is unlikely that a different choice of non-inferiority margin or statistical analytical approach to the non-inferiority results, or a longer waiting time would change the final conclusion; namely, that the HATCP has limited value as a structural graft in unfixed CLOs in paediatric patients with PPV.

References

1. Blitz NM, Stabile RJ, Giorgini RJ, DiDomenico L a. Flexible Pediatric and Adolescent Pes Planovalgus: Conservative and Surgical Treatment Options. Clin Podiatr Med Surg. 2010;27:59–77.
2. John S, Child BJ, Hix J, Maskill M, Bowers C, Catanzariti AR, et al. A retrospective analysis of anterior calcaneal osteotomy with allogenic bone graft. J Foot Ankle Surg. 2010 Jan;49(4):375–9.
3. Dimitriou R, Mataliotakis GI, Angoules AG, Kanakaris NK, Giannoudis P V. Complications following autologous bone graft harvesting from the iliac crest and using the RIA: A systematic review. Injury. Elsevier Ltd; 2011;42:S3–15.
4. Myeroff C, Archdeacon M. Autogenous Bone Graft : Donor Sites and Techniques. J Bone Jt Surg. Elsevier; 2011 Dec 7;93(23):2227–36.

5. Kurien T, Pearson RG, Scammell BE. Bone graft substitutes currently available in orthopaedic practice: The evidence for their use. Bone Joint J. 2013 May;95-B(5):583–97.
6. Lobo SE, Arinzeh TL. Biphasic calcium phosphate ceramics for bone regeneration and tissue engineering applications. Materials (Basel). 2010;3(2):815–26.
7. Campana V, Milano G, Pagano E, Barba M, Cicione C, Salonna G, et al. Bone substitutes in orthopaedic surgery: from basic science to clinical practice. J Mater Sci Mater Med. Springer; 2014 Oct;25(10):2445–61.
8. Lindaman LM. Bone healing in children. Clin Podiatr Med Surg. 2001 Jan;18(1):97–108.
9. Templin D, Jones K, Weiner DS. The Incorporation of Allogeneic and Autogenous Bone Graft in Healing of Lateral Column Lengthening of the Calcaneus. J Foot Ankle Surg. 2008;47(4):283–7.
10. Vining NC, Warme WJ, Mosca VS. Comparison of structural bone autografts and allografts in pediatric foot surgery. J Pediatr Orthop. 2012;32(7):719–23.
11. Anand A, Feldman DS, Patel RJ, Lehman WB, van Bosse HJP, Badra MI, et al. Interobserver and intraobserver reliability of radiographic evidence of bone healing at osteotomy sites. J Pediatr Orthop B. 2006;15(4):271–2.
12. Balakumar B, Babu S, Varma HK, Madhuri V. Triphasic ceramic scaffold in paediatric and adolescent bone defects. J Pediatr Orthop B. 2014;23:187–95.
13. Lee IH, Chung CY, Lee KM, Kwon S-S, Moon SY, Jung KJ, et al. Incidence and risk factors of allograft bone failure after calcaneal lengthening. Clin Orthop Relat Res. 2015 May;473(5):1765–74.
14. Martinkevich P, Rahbek O, Moller-Madsen B, Soballe K, Stilling M. Precise and feasible measurements of lateral calcaneal lengthening osteotomies by radiostereometric analysis in cadaver feet. Bone Jt Res. 2015;4(5):78–83.
15. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of Noninferiority. 2012;
16. Mosca VS. Calcaneal Lengthening for Valgus Deformity of the Hindfoot. J Bone Jt Surg Am. 1995;
17. Valstar ER, Gill R, Ryd L, Flivik G, Börlin N, Kärrholm J. Guidelines for standardization of radiostereometry (RSA) of implants. Acta Orthop. 2005;76(4):563–72.
18. Ranstam J, Ryd L, Onsten I. Accurate accuracy assessment: review of basic principles. Acta Orthop

Scand. 2000 Mar;71(1):106–8.

19. Davids JR, Gibson TW, Pugh LI. Quantitative segmental analysis of weight-bearing radiographs of the foot and ankle for children: normal alignment. - PubMed - NCBI. J Pediatr Orthop. 2005;25(6):769–76.

20. Cavanagh PR, Rodgers MM. The arch index: A useful measure from footprints. J Biomech. 1987 Jan;20(5):547–51.

21. Martinkevich P, Møller-Madsen B, Gottliebsen M, Kjeldgaard Pedersen L, Rahbek O. Validation of the translated Oxford ankle foot questionnaire in 82 Danish children aged between five and 16 years. Bone Jt J. 2015;97B(3):420–6.

22. Sealed Envelope Ltd 2012 [Internet]. [cited 2015 Nov 11]. Available from: www.sealedenvelope.com/power/continuous-noninferior

23. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004 Aug;240(2):205–13.

24. Ploghaus A, Narain C, Beckmann CF, Clare S, Bantick S, Wise R, et al. Exacerbation of pain by anxiety is associated with activity in a hippocampal network. J Neurosci. 2001 Dec 15;21(24):9896–903.

25. Clarke A, Flowers MJ, Davies AG, Fernandes J, Jones S. Morbidity associated with anterior iliac crest bone graft harvesting in children undergoing orthopaedic surgery: a prospective review. J Child Orthop. Springer Berlin Heidelberg; 2015;9(5):411–6.

26. Scholl R, Bekker A, Babu R. Neuroendocrine and Immune Responses to Surgery. Internet J Anesthesiol. 2012;30(3).

27. Kozlowski LJ, Kost-Byerly S, Colantuoni E, Thompson CB, Vasquenza KJ, Rothman SK, et al. Pain prevalence, intensity, assessment and management in a hospitalized pediatric population. Pain Manag Nurs. 2014 Mar;15(1):22–35.

28. Rabello F, Souza C, Farina Jr J. Update on hypertrophic scar treatment. Clinics. Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo; 2014 Aug 12;69(8):565–73.

29. Friedmann A, Gissel K, Konermann A, Götz W. Tissue reactions after simultaneous alveolar ridge

augmentation with biphasic calcium phosphate and implant insertion—histological and immunohistochemical evaluation in humans. Clin Oral Investig. 2014;

30. Becker K, Heilmann C, Peters G. Coagulase-Negative Staphylococci. Clin Microbiol Rev. 2014 Oct 2;27(4):870–926.

31. Pocock SJ. EDUCATION & DEBATE When to stop a clinical trial. BMJ. 1992;305:235–40.