UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH SCIENCES

PhD Thesis

Rudi Neergaard Hansen

Patellofemoral pain: Comparable effectiveness of exercises, effect modifiers, and agreement of visual assessments

Supervisors: Marius Henriksen, Peter Magnusson, Christian Couppé
This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen at Dec. 21st, 2022.

Patellofemoral pain: Comparable effectiveness of exercises, effect modifiers, and agreement of visual assessments

Name of department: Dep. of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital

Author(s): Rudi Neergaard Hansen

Topic description: The main topic of this thesis is the treatment of patients with patellofemoral pain. It includes three studies that each contributes to different aspects: The comparable effectiveness of two training modalities, the exploration of potential effect modifiers, and the reproducibility of visual assessments of knee alignment.

Supervisors: **Marius Henriksen** (main supervisor), Professor, PhD, PT
The Parker Institute, Bispebjerg and Frederiksberg Hospital

Peter Magnusson, Dr. med., PT
Dep. of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital
Institute of Sports Medicine Copenhagen, Bispebjerg and Frederiksberg Hospital
Center for Healthy Aging, Faculty of Health and medical Sciences, University of Copenhagen

Christian Couppé, Associate Professor, PhD, PT
Dep. of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital
Institute of Sports Medicine Copenhagen, Bispebjerg and Frederiksberg Hospital

Assessment Committee: **Inger Mechlenburg**, ph.d., Professor (opponent)
Department of Clinical Medicine,
Aarhus University

Bradley Neal, ph.d., Lecturer (opponent)
School of Sport, Rehabilitation and Exercise Sciences,
University of Essex

Thomas Quaade Bandholm, ph.d., Professor (chairman)
The Department of Clinical Medicine,
University of Copenhagen.

Submitted on: 21 December 2022

Number of study units: 3
Number of characters: 83764
Number of words: 12732

This thesis is based on the following studies:

Study 1 (appended in page 162): Hansen R, Brushøj C, Rathleff MS, Magnusson SP, Henriksen M. Quadriceps or hip exercises for patellofemoral pain? A randomized controlled equivalence trial (submitted to British Journal of Sports Medicine).

Study 2 (appended in page 186): Hansen R, Brushøj C, Rathleff MS, Magnusson SP, Henriksen M. Differential effects of quadriceps and hip muscle exercise programs for patellofemoral pain: A secondary effect modifier analysis of a randomized trial (ready for submission).

Study 3 (appended in page 207): Hansen R, Lundgaard-Nielsen M, Henriksen M. Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study. PeerJ. 2021;9:e12203.

Preface and acknowledgement

I have been privileged by three exceptional supervisors during this PhD project. You have all led by great example, and I have learned so much from your different perspectives and styles of mentorship. First of all, thanks to my primary supervisor, professor Marius Henriksen, for endless and skillful support. You have given me the autonomy to do research that I am excited about, the encouragement to direct those projects, and the mentorship to help me succeed. I admire your academic competences, and if I can absorb just a bit of your knowledge, I will grow to be a decent researcher. Further, I would like to express my appreciation to my co-supervisors, professor Peter Magnusson and associate professor Christian Couppé, for support, inspirational discussions and for sharing your extensive knowledge. You have taken me under your wings and supported me since day one, and I am deeply grateful.

To my dearest colleagues Camilla Cullum, Mathilde Lundgaard-Nielsen, Robert Herzog, Christina Eddy, Sebastian Åxman, Jakob Kristiansen, Rikke Høffner, Tone Heiberg, and Christina Kaae: Thanks for your never-ending support and for your major contributions to the studies in this PhD. Without you, this work would have never been possible. Also, a big thanks to the medical doctors at the Institute of Sports Medicine Copenhagen for helping me include the many participants.

Thanks to the head of the Department of Physical and Occupational Therapy at Frederiksberg/Bispebjerg Hospital Naja Henriksen and to my nearest leader Birgitte Randeris for the support and for believing in the project. We co-developed new ways of being a PhD student, that can hopefully inspire future researchers.

Thanks to the co-authors of the studies included in this thesis: Michael Skovdal Rathleff for sharing your boundless expertise in the field of patellofemoral pain and for your priceless contributions to the manuscripts and Christoffer Brushøj for your competent feedback and input to the design of the studies and the manuscripts.

A special thanks to all the participants who struggled with exercises, compliance, pain and much more. You have all contributed to the research puzzle of patellofemoral pain, although it seems that we still have a long way to go.

Last but not least, thanks to Marie, my wife, and Aksel and Agnes, my kids, for patience, support, understanding, and cheering. You have been a great source of encouragement, and I could not have done this without your support.

Table of contents

Summary (English)

Introduction: International consensus based on current evidence advocate the use of resistance training for the lower extremities for the treatment of patients with patellofemoral pain (PFP). But direct comparisons of exercise modalities are few, sample sizes are small, and intervention- and follow-up periods have been short. Collectively, this challenges the choice of the most effective treatment for the patient group. Moreover, the PFP population is very heterogenous and a 'one-size-fits-all' approach presumably is not optimal. We need to identify patient characteristics that can predict a superior outcome from one specific treatment over the other. One of these potential characteristics is functional alignment of the knee. However, the reproducibility of assessing functional knee alignment in patients with PFP in the clinic has not been established.

Purpose: The primary purpose of this PhD project was to investigate the comparable effectiveness of two commonly prescribed exercise protocols on symptoms and physical function for patients with PFP. Further, we aimed to assess which patient characteristics that predict a better outcome of one of the exercise protocols over the other. As a final purpose, we wanted to investigate the intra- and interrater agreement for visual assessments of knee alignment during a single-leg squat (SLS) and a forward lunge (FL).

Methods: To address the purposes set above, three studies were completed. First, a randomized clinical equivalence trial exploring the comparable effectiveness of a hip focused vs a quadriceps focused exercise program was conducted (Study 1). Secondly, a post-hoc analysis exploring the interaction between certain pre-specified patient characteristics and group allocation was performed (Study 2). Finally, we conducted a cross-sectional agreement study to establish the reproducibility of visual assessments of knee alignment (Study 3).

In Study 1, we included 200 participants diagnosed with PFP. They were randomly allocated to either quadriceps focused exercises (QE), including squat, lunge, and knee extension or hip muscle focused exercises (HE), including hip abductions, clam-shell, and hip extensions. Upon an instructional consultation, participants did their exercises three times per week at home with monthly clinical visits at the clinic. The primary outcome was change from baseline in the Anterior Knee Pain Scale (AKPS) at week 12. Secondary outcomes included change in the Knee injury and Osteoarthritis Outcome Score (KOOS), isometric muscle strength, Dynamic Assessment of Pain (DAP), Pain Self-Efficacy Questionnaire, the 3-level version of EuroQoL 5 dimensions (EQ-5D-3L) Questionnaire, and The Transition Questionnaire of global perceived effect on overall health, pain, and function at week 12. Change from baseline in questionnaire data was re-assessed at week 26.

In Study 2, we performed a post-hoc analysis on the population from the main trial. Primary outcome was the AKPS at week 12, and the candidate baseline characteristics encompassed a range of self-reported information as well as clinical observations. The analyses focused on the interaction between presence/absence of the patient characteristics and group allocation (QE vs. HE) at each time point (week 12 and 26).

In Study 3, we included the first 60 participants from the main study. A video was recorded using a tablet from an anterior view of participants performing the SLS and FL. The investigator scored the movement as observed clinically according to preset criteria (valgus, varus, or no malalignment). At least one week later, the investigator did another scoring based on the recorded video and another investigator repeated the scoring independently. Cohen's weighted kappa statistics was used to assess the intra- and interrater agreement.

Findings: We included 200 participants for Study 1 and 2 (mean age 27.2 years, SD 6.4); 60 of these were included for Study 3. Mean changes in AKPS questionnaire score from baseline to week 12 were 7.5 (SE ±0.8) for QE and 7.2 (SE ±0.8) for HE. The 95% CI of the group difference in change in AKPS questionnaire from baseline to week 12 was within the predefined equivalence margin of±8 points; $p<0.0001$ for equivalence. The key secondary outcomes were all within the predefined criteria for equivalence. Further, we found a significant subgroup difference at week 12 in favor of QE among participants with pain catastrophizing at baseline compared to those without pain catastrophizing (subgroup difference 8.3 AKPS points (95% CI 1.6 to 15.0; $p=0.016$). At 26 weeks, participants with a baseline BMI ≥25 m/kg^2 benefitted from QE compared to those with BMI <25 m/kg^2 with a subgroup difference of 11.1 AKPS points (95% CI 4.8 to 17.4; $p=0.001$). In contrast, participants with severe baseline knee pain seemed to benefit from HE when compared to those with mild-moderate baseline knee pain with a subgroup difference of -9.1 AKPS points (95% CI -15.7 to -2.6; $p=0.006$). Lastly, we found that Kappa values for the intrarater agreement when visually assessing dynamic knee alignment ranged from 0.58 to 0.70, i.e., moderate to good agreement, whereas the interrater agreement ranged from 0.22 ($p=0.08$) to 0.50 ($p<0.0001$), i.e., fair to moderate agreement.

Conclusion: We found that an exercise program that focused on either quadriceps or hip muscles provided equivalent improvements in symptoms and function in the short (12 weeks) and medium term (26 weeks). Further, we found that participants with pain catastrophizing or a high BMI benefitted more from quadriceps exercises than from hip exercises, whereas hip exercises were better for patients with more severe knee pain at baseline. Lastly, we found moderate to good intrarater agreement visually assessing dynamic knee alignment during SLS and FL, whereas the interrater agreement ranged from fair to moderate.

Resumé (Dansk)

Introduktion: I international forskning er der konsensus om og evidens for at anbefale styrketræning til underekstremiteterne til behandling af patienter med patellofemorale smerter (PFP). Dog har kun få studier sammenlignet forskellige træningsmodaliteter, sample sizes har generelt været små og interventions- og follow-up-perioder har været korte. Dette gør det udfordrende at vælge den mest effektive behandling til patientgruppen. Desuden er PFP-populationen meget heterogen, hvorfor en 'one-size-fits-all' tilgang formodentlig ikke er optimal. Vi er derfor nødt til at identificere patientkarakteristika, der kan forudsige et bedre resultat af den ene behandling frem for den anden. En af disse potentielle patientkarakteristika er dynamisk alignment af knæet. Reproducerbarheden af at vurdere knæalignment hos patienter med PFP i klinikken er dog ikke blevet undersøgt.

Formål: Det primære formål med dette ph.d.-projekt var at sammenligne effekten af to ofte anvendte træningsprotokoller målt på symptomer og fysisk funktion hos patienter med PFP. Herudover ville vi undersøge, hvilke patientkarakteristika der kunne forudsige et bedre resultat af den ene slags træning frem for den anden. Som et sidste formål ønskede vi at undersøge intra- og interterrater agreement når terapeuter vurderer dynamisk knæalignment under en etbenssquat (SLS) og en lunge (FL).

Metoder: For at undersøge formålene angivet ovenfor blev tre forskningsstudier gennemført. Vi udførte et randomiseret klinisk ækvivalensstudie, der undersøgte effekten af et hoftefokuseret versus et quadricepsfokuseret træningsprogram (Studie 1). Herudover lavede vi en post-hoc-analyse, der undersøgte interaktionen mellem de præspecificerede patientkarakteristika og interventionsgruppe (Studie 2). Endelig gennemførte vi en tværsnitsundersøgelse for at fastslå reproducerbarheden af visuel vurdering af dynamisk knæalignment (Studie 3).

I Studie 1 inkluderede vi 200 deltagere diagnosticeret med PFP. De blev tilfældigt allokeret til enten quadriceps øvelser (QE), inklusive squat, lunge og knæstræk eller hofteøvelser (HE), inklusive hofteabduktioner, muslingeøvelsen og hofteekstensioner. Efter en instruktion lavede deltagerne deres øvelser tre gange om ugen derhjemme med månedlige kontrolbesøg i klinikken. Det primære effektmål var ændring fra baseline i Anterior Knee Pain Scale (AKPS) ved uge 12. Sekundære effektmål var ændring i Knee injury and Osteoarthritis Outcome Score (KOOS), isometrisk muskelstyrke, Dynamic Assessment of Pain (DAP), Pain Self-Efficacy spørgeskema, EuroQoL 5 dimensions (EQ-5D-3L) spørgeskema og Transition Questionnaire of global perceived effect on overall health, pain, and function ved uge 12. Ændring fra baseline i spørgeskemaundersøgelserne blev revurderet i uge 26.

I Studie 2 udførte vi en post-hoc analyse af populationen fra hovedstudiet. Det primære effektmål var AKPS ved uge 12, og patientkarakteristika omfattede en række selvrapporterede

oplysninger samt kliniske observationer. Analyserne fokuserede på interaktionen mellem tilstedeværelse/fravær af patientkarakteristika og interventionsgruppe (QE vs. HE) ved uge 12 og 26.

I Studie 3 inkluderede vi de første 60 deltagere fra hovedundersøgelsen. Deltagerne udførte en SLS og FL, og undersøgeren vurderede og scorede bevægelsen i knæet i henhold til fastsatte kriterier (valgus, varus eller ingen fejlstilling). Der blev samtidig optaget en video af bevægelserne. Mindst én uge senere lavede undersøgeren en ny vurdering baseret på den optagede video, og en anden undersøger gentog vurderingen uafhængigt. Cohens kappa blev brugt til at analysere intra- og interrater agreement.

Resultater: Vi inkluderede 200 deltagere til Studie 1 og 2 (gennemsnitsalder 27,2 år (SD 6,4)); 60 af disse blev inkluderet i Studie 3. Gennemsnitlige ændringer i AKPS-score fra baseline til uge 12 var 7,5 (SE ±0,8) for QE og 7,2 (SE ±0,8) for HE. 95% konfidensintervaller for ændring i AKPS-score fra baseline til uge 12 lå inden for den foruddefinerede ækvivalensmargin på ±8 point; p<0,0001. De sekundære effektmål lå også alle inden for de foruddefinerede kriterier for ækvivalens. Ydermere fandt vi en signifikant forskel mellem subgrupper i uge 12 til fordel for QE blandt deltagere med smertekatastrofering ved baseline sammenlignet med dem uden smertekatastrofering (forskel på 8,3 AKPS-point (95 %CI 1,6 til 15,0; p=0,016). Efter 26 uger havde deltagere med en baseline BMI ≥25 m/kg2 fordel af QE sammenlignet med dem med BMI <25 m/kg2 med en forskel på 11,1 AKPS-point (95 %CI 4,8 til 17,4; p=0,001). Deltagere med svære knæsmerter ved baseline syntes at have gavn af HE sammenlignet med dem med milde til moderate knæsmerter med en subgruppe forskel på -9,1 AKPS-point (95 %CI -15,7 til -2,6; p=0,006). Endelig fandt vi, at Kappa værdier for intrarater-agreement ved visuel vurdering af dynamisk knæalignment varierede fra 0,58 til 0,70 dvs. moderat til god overensstemmelse, hvorimod interrater-agreement varierede fra 0,22 (p=0,08) til 0,50 (p<0,0001) dvs. rimelig til moderat overensstemmelse.

Konklusion: Vi fandt ud af, at quadriceps- og hofteøvelser gav ækvivalente forbedringer i symptomer og funktion på kort (12 uger) og mellemlang sigt (26 uger). Yderligere fandt vi, at deltagere med smertekatastrofering eller et højt BMI havde mere gavn af quadricepsøvelser end af hofteøvelser, hvorimod hofteøvelser var bedre for patienter med mere alvorlige knæsmerter ved baseline. Endelig fandt vi moderat til god intrarater-agreement ved visuel vurdering af dynamisk knæalignment under SLS og FL, hvorimod interrater-agreement varierede fra rimelig til moderat.

Introduction

Patellofemoral pain (PFP) is considered one of the most common musculoskeletal disorders (1, 2). PFP is characterized by significant pain behind or around the patella (kneecap) aggravated by squatting, stair climbing, running, and prolonged sitting with the knees bend (3). Patients are often young, symptom duration is long, and the impact on activity levels, participation in sports, and ultimately quality of life, is substantial (4). The prevalence is high, and the prognosis is worse than for other knee disorders (5). This implies high burdens on the individual and on society (6, 7).

The pathogenesis of PFP is probably multifactorial and not fully understood (8). Therefore, treatment is nonspecific and non-curative and mainly aims at relieving pain and increasing function (9). Exercise is well documented as a treatment for pain and physical impairment in patients with PFP and is recommended as a core element of treatment (10, 11). The latest consensus statements recommend inclusion of resistance exercises for the quadriceps and hip muscles (10, 11). Traditionally, quadriceps based exercise programs have been recognized as cornerstones in PFP rehabilitation (12). It is theorized that patella instability and/or maltracking causes the retro- and peripatellar pain (13), and that strength training of the quadriceps muscle improves stability and alignment, thereby relieving the pain. The rationale for applying hip strengthening is to address the functional stability and alignment of the lower extremity (14). Moreover, an analgesic effect of exercising a non-painful limb has been reported in chronic pain conditions, whereas exercising the painful limb, e.g., in knee osteoarthritis, is associated with hyperalgesia (15, 16). Direct comparisons of exercise protocols are few (11, 17), with short intervention follow-up periods, and with insufficient sample sizes (18, 19). Moreover, the PFP population is very heterogeneous (9) presumably making "one-size-fits-all"-approaches to exercise prescription suboptimal. It is therefore unclear which patients that will benefit most from different types of treatment, including quadriceps and hip exercises (20, 21). Collectively, this challenges the choice of the most appropriate treatment for the individual patient and may also explain the lack of differences in direct comparisons of different exercise types (22). There is a need for large high-quality studies of comparative effectiveness of quadriceps and hip muscle exercises for PFP and for identification of subgroups that respond differentially to different exercise regimens.

Aims of the thesis

The overall aim of this thesis was to compare the effectiveness of a focused Quadriceps Exercise program (QE) vs a focused Hip Exercise program (HE) and to identify which patients that can expect to gain more from one exercise program than the other. This was addressed in three studies with the following individual aims:

1. To assess effectiveness equivalence between QE and HE on symptoms and function in a population of PFP-patients (Study 1)
2. To identify contextual factors that modify the observed treatment effect of the two treatments (QE vs. HE) across patient subgroups, i.e., assess whether the treatment effect is modified by the value of a variable assessed at baseline (Study 2)
3. As one of the contextual factors that was planned to be assessed in Study 2 was knee alignment during a single leg squat (SLS) and a forward lunge (FL), the third aim was to determine the intrarater and interrater agreement of a subjective visual scoring of dynamic knee alignment during SLS and FL (Study 3).

Hypotheses

1. QE and HE have equivalent effectiveness on self-reported function and symptoms after 12 weeks of treatment in patients with PFP (Study 1)
2. One or more baseline characteristic can predict a superior outcome of either QE or HE (Study 2)
3. Visual assessments of dynamic knee joint alignment during a FL and a SLS performed by patients with PFP can be done reliably regarding both intrarater and interrater agreement (Study 3)

Background

Epidemiology

More than half of the Danish population experience musculoskeletal pain at a regular basis and the number is increasing (23). Each year there are 10 million consultations in general practice due to musculoskeletal pain. Overall, this type of pain costs the Danish society DKK 16 billion (EUR 2.15 billion) every year (24). PFP is considered one of the most common musculoskeletal disorders seen in the clinic of general practitioners and physiotherapists (25, 26). A lifetime prevalence as high as 45% among females have been reported (27), and the annual prevalence is 23% in mixed populations (2). Point prevalence is 7.2% in a general adolescent population and 16.3% in female only adolescents (2). Incidence rates in military recruits is reported to range between 9.7 and 571.4 per 1000 person-years, in amateur runners 1080.5 per 1000 person-years and in amateur athletes between 5.1% and 14.9% per season (2). The incidence of PFP during a long-distance running race range between 5.9% and 25.6% (28, 29). With regards to gender specific differences, females are twice as likely to experience PFP as males (30).

Collectively, the data highlight, that PFP is widespread in the population. But variations in the reporting of prevalence and incidence is apparent. The reason for this is likely differences in populations, definitions, and diagnostic criteria between studies.

Etiology

The etiology of PFP is unknown, but PFP is described as nontraumatic in nature and presents with pain behind and/or around the kneecap on activities that load the joint such as squatting, running, climbing, and descending stairs (3, 31). Traditionally a biomechanical model has been applied when describing the causes of PFP. In this model, the disorder is linked to over- or misuse of the patellofemoral joint leading to joint stress, pain and even osteoarthritis (OA) (32-34). Several contributing factors to the pathophysiology of PFP in this model have been proposed including knee malalignment (35, 36), patella maltracking (37, 38), imbalances in muscle strength and architecture (39, 40), and foot posture (41). The presentation of biomechanical deficits in people with PFP is commonly reported with the most reported being a larger Q-angle, larger sulcus angle, larger patellar tilt angle, less hip abduction strength, and less knee extension strength compared to pain free subjects (42). Theoretically, the maltracking increases patellofemoral joint stress, initiating nociceptive firing from the subchondral bone or soft tissue (33).

Some suggest that PFP and patellofemoral OA exist along a continuum of disease (32, 43, 44). No studies, however, have prospectively verified this relationship (32). A retrospective case-control study observed that patients under-going arthroplasty for patellofemoral OA were more than twice as likely to report having had PFP as an adolescent than patients undergoing arthroplasty for isolated tibiofemoral OA (45). In another study including adults aged 40 years and older with chronic PFP, radiographic patellofemoral OA was present in 69% of people, and was more prevalent than tibiofemoral OA (45%) (46). This infers a relationship that is strengthened by the similarities in symptoms and impairments between PFP and patellofemoral OA (32). In line with this, studies indicate that the pain of PFP may originate from the subchondral bone areas in the joint between the patella and the femur. A study with PET/CT using the tracer [^{18}F]sodium fluoride showed increased bone remodeling in the subchondral bone structures of the knee joint in people with PFP and a correlation between increasing bone remodeling and increasing pain intensity (47). The increased activity in the study was thought to occur due to increased mechanical load in the joint possibly related to poor tracking of the patella and can be seen as an expression of a healing process and changes which are also seen in early OA.

Whereas the biomechanics (wear and tear) is one hypothetical theory in the development of PFP, another proposed mechanism is pain sensitization. Pain sensitization has been defined by the International Association for the Study of Pain as "*increased responsiveness of nociceptive neurons in the central nervous system"* (48). In patients with persistent PFP, local and widespread sensitization has been reported, indicating altered pain processing (49-52). Widespread sensitization is common in other knee disorders such as knee OA with spreading of pain beyond the local painful area (53). It has been proposed that tissue stress can lead to sensitization of peripheral nociceptors (54), hence increasing pain. But pain can persist even when patients modify or reduce physical activity (5). Therefore, the phenomenon may not be driven nor maintained entirely by load. Characteristics such as catastrophizing thoughts, fear-of-movement and pain self-efficacy have been associated with pain and disability and recognized as barriers to recovery (55). A psychologically informed intervention targeting pain-related fear and pain catastrophizing, did not, however, induce any significant group difference compared to a control intervention in a recently conducted randomized controlled trial (RCT) (56). And in a mediation analysis of the plausible mechanisms of the treatment effect of exercises, changes in anxiety, kinesiophobia, or pain catastrophizing did not mediate the effect of a training intervention (57). Like with the biomechanical paradigm, methodological limitations make it challenging to causally link psychological features with the etiology of PFP. The evidence is

based on cross sectional and retrospective analyses and any causal relationship with the development of PFP has not been established (42). Pain catastrophizing, fear-avoidance beliefs, and pain self-efficacy are considered clinically important, and it is emphasized that researchers should prioritize investigating the psychological features of PFP in the future (58).

Assessment

As described, the etiology of PFP is considered multifactorial but largely unknown and the diagnosis is made on a "diagnosis by exclusion" basis in the absence of other identifiable pathology such as meniscal injury, tendinopathy, bursitis or apophysitis (3, 59, 60). Therefore, the accuracy of clinical and functional tests for diagnosing PFP is low (59). A systematic review concluded that the presence of pain during a squatting test is most accurate in diagnosing PFP, although it could not be considered sufficient as a basis for PFP diagnosis (59). Combinations of functional assessment tests including e.g., pain during muscle contraction, pain during squatting, and pain during palpation yielded no better diagnostic value for PFPS than did the squat test per se (61). Methodological limitations make it difficult to draw any decisive conclusions on which diagnostic tests are the best, including insufficient blinding, assessing healthy participants vs. PFP patients, and lack of a solid gold standard as a reference test (60).

In the clinic a thorough history taking is recommended. Patients typically present with insidious onset of pain unrelated to trauma and with gradual progression of pain (3). Patients describe pain behind or around the patella, often provoked by activities that stress the patellofemoral joint, such as stairclimbing, squatting, kneeling, and prolonged sitting with knee flexed (31). Further, a clinical exam including visual assessment of alignment and movement quality will highlight any biomechanical variations or altered neuromuscular function. Kinematic alterations like knee valgus, pelvic drop, femoral anteversion, and high Q-angle has been linked to PFP (31, 38, 62, 63) but are not independent risk factors (64). The step-down test, single leg squat, and forward lunge tests are methods that can be reliably used in the biomechanical examination (65, 66). However, most evaluations have been performed using time consuming and sophisticated assessment equipment, e.g., 2- or 3-Dimensional motion capture systems (67, 68). Therapists rely on visual assessments in the clinic as they often do not have access to the equipment or time required for complex biomechanical analysis. Most studies have included only pain free participants (69, 70), and therefore the reliability of visual assessment of knee alignment in PFP patients in the clinic is unclear.

As stated in the 'Etiology' section psychological factors related to patients' cognitive appraisal of their pain experience may play a role in the development and persistence of PFP. As

in other pain disorders fear of pain and pain anxiety may lead to avoidance behaviors, which can result in disuse, disability, and depression (58). Change in fear-avoidance beliefs has been reported to predict function and pain outcomes in PFP (71). It is therefore recommended that clinicians consider pain catastrophizing, pain self-efficacy, and fear-avoidance beliefs when undertaking a clinical examination (58). Further, pain sensitization should be considered and examined. Clinically a variety of diagnostic markers are used for the assessment of pain sensitization including questionnaires, simple sensory testing, and mapping of areas with sensory abnormalities (72).

Treatment

Like for other chronic musculoskeletal pain conditions, there are many different treatment modalities proposed to be beneficial for patients with PFP. These include patellar taping (73), dry needling (74), photobiomodulation (75), whole-body vibration (76), and diacutaneous fibrolysis (77), just to mention a few. The many proposed modalities have limited effect and reflect that the underlying causes of PFP are unknown (3). It has been repeatedly emphasized that active rehabilitation comprising strengthening exercises for the hip, the knee or both the hip and knee should be the cornerstone of treatment, best combined with education in pain and guidance on how to manage pain without stopping exercise (10, 78, 79). A wait-and-see approach is not considered appropriate, and therefore people with PFP should be referred for active rehabilitation or combined interventions, i.e., exercise plus another intervention (80).

Studies including direct comparisons of the separate exercise protocols are few and intervention durations and follow-up periods have been short (17, 19, 80-84). Whereas some potential benefits of exercise may be acute, e.g., exercise induced pain relief (85), others will likely need longer time to arise. Muscle hypertrophy and structural adaptations of soft tissue are usually minimal during the first 4 weeks of a resistance training program (86), implying that short term rehabilitation periods may not be sufficient to benefit fully from training. Concerning the follow-up periods, there is a dearth of high-quality exercise therapy trials beyond a short-term follow up which highlights the disparity between reported symptom persistence and follow up duration (80). Collectively, this challenges the choice of the most appropriate treatment and may also explain the variation in effect in clinical practice (13).

Quadriceps focused exercises

Quadriceps based exercise programs have been recognized as cornerstones in PFP rehabilitation (12), but the mechanisms by which quadriceps exercises affect pain and function are not clear.

Patients with PFP demonstrate less knee extensor strength than controls (87, 88), and it is theorized that patella instability and/or maltracking causes the retro- and peripatellar pain (13). In the biological model, a focus is on the biomechanical changes induced by exercises, i.e., increasing muscle strength, altering neuromuscular function, increasing flexibility, improving stability, optimizing alignment etc. (57, 89). Strength training of the quadriceps muscle proposedly improves knee stability and alignment, thereby relieving the pain (13). Another proposed mechanism of the effect of exercises for PFP is the analgesic effect. It is well known that exercise produces analgesic effects in chronic pain patients through numerous biopsychosocial mechanisms (85, 90). However, exercising painful joints have also been reported to have the opposite effect in chronic pain (91, 92).

Hip focused exercise

There is growing evidence that impaired hip muscle strength and function is associated with PFP (93, 94). This often is observed as excessive hip adduction and/or internal rotation during activities such as running and stair climbing (95, 96) .The rationale for applying hip strengthening is that the patellofemoral joint may be stressed by the excessive dynamic valgus and medial femoral rotation during weight bearing activities caused by poor hip joint control due to muscle weakness (97, 98). Therefore, strengthening the hip abductors and external rotators potentially modifies the biomechanics and may be beneficial in the rehabilitation process. An increasing quantity of literature supports the implementation of hip strengthening exercises in the rehabilitation process (81, 99-101).

Another proposed mechanism concerns the analgetic effect of exercising a distant non-painful joint that may be associated with exercise induced hypoalgesia. In patients with knee OA, exercising a distant body region induced decreased pain sensitivity, whereas exercising the knees did not (15, 16).

Treatment effect modifiers

It can be hypothesized that certain patient characteristics may predict outcome success of either a hip training program or a training program that focus on quadriceps training, but this remains to be shown. Identifying subgroups of individuals who may benefit more from one treatment than the other and potential treatment effect modifiers is an important goal in health research (102, 103). Whereas a treatment might be effective for some, it may be ineffective or even harmful for others. Therefore, increased focus has been put on modifiers of treatment effect on outcome in clinical trials. Hereby, individual differences between patients can be acknowledged and may

inform identification of subgroups who are likely to benefit from one type of treatment over the other (104).

The PFP population is described as very heterogenous in terms of e.g., age, activity level, biomechanical and psychosocial factors, and symptoms (3, 105, 106). Substantial efforts have been made to identify which patients or patient characteristics that predict outcome in conservative PFP management (6, 20, 107). Several factors have been linked to a poor outcome with longer duration of symptoms, older age, lower function, bilateral symptoms, and number of pain sites at baseline being the most reported (20, 21, 108-111). However, although prognostic factors help predict the likelihood of an outcome within a certain time period, they cannot predict which specific treatment is the best (20, 112). There is a paucity of literature on effect modifiers and therefore, adequately designed randomized trials are needed.

Outcome assessment tools

When evaluating the treatment response in patients with PFP, valid, reliable, responsive, and specific assessment tools are of outmost importance – both for the clinician to effectively monitor treatment response and for the researcher. Several outcome measures have been used in clinical trials of interventions for patellofemoral pain, but since PFP is a nonspecific pain disorder with no signs of tissue specific abnormalities, no objective measure is appropriate. Therefore outcomes are most often evaluated by patient reported outcome measures (PROMs) that address function, pain and disability (113). PROMs are essential for evaluating treatment impact of many conditions including pain syndromes (114), but there are limitations and biases one needs to be aware of. PROMs may be affected by mood, expectations, time, and interactions with the therapist. Further, patients often tend to overestimate the benefits of a treatment (115).

As pain is one of the main features of PFP, several trials have used a pain scale to assess outcome, e.g., a visual analog scale (VAS), numeric rating scale (NRS), or a verbal rating scale (VRS). Furthermore, several PROMs have been developed to assess the level of symptoms and disability of persons affected by knee pain (116) and specifically for PFP (117, 118). Lastly, measures of global ratings of improvement, participant satisfaction, physical activity level, and clinical evaluation have also been frequently used in PFP trials (116). Several secondary outcomes measuring physiologic and psychologic aspects not directly related to PFP, e.g., muscle strength, mobility, knee alignment, and pain self-efficacy, are also frequently used and give additional information on the effects and characteristics of the disorder (8).

The Anterior Knee Pain Scale (AKPS) and VAS for usual or worst pain are found to be the most reliable, valid, and responsive and are therefore recommended for clinical trials in PFP

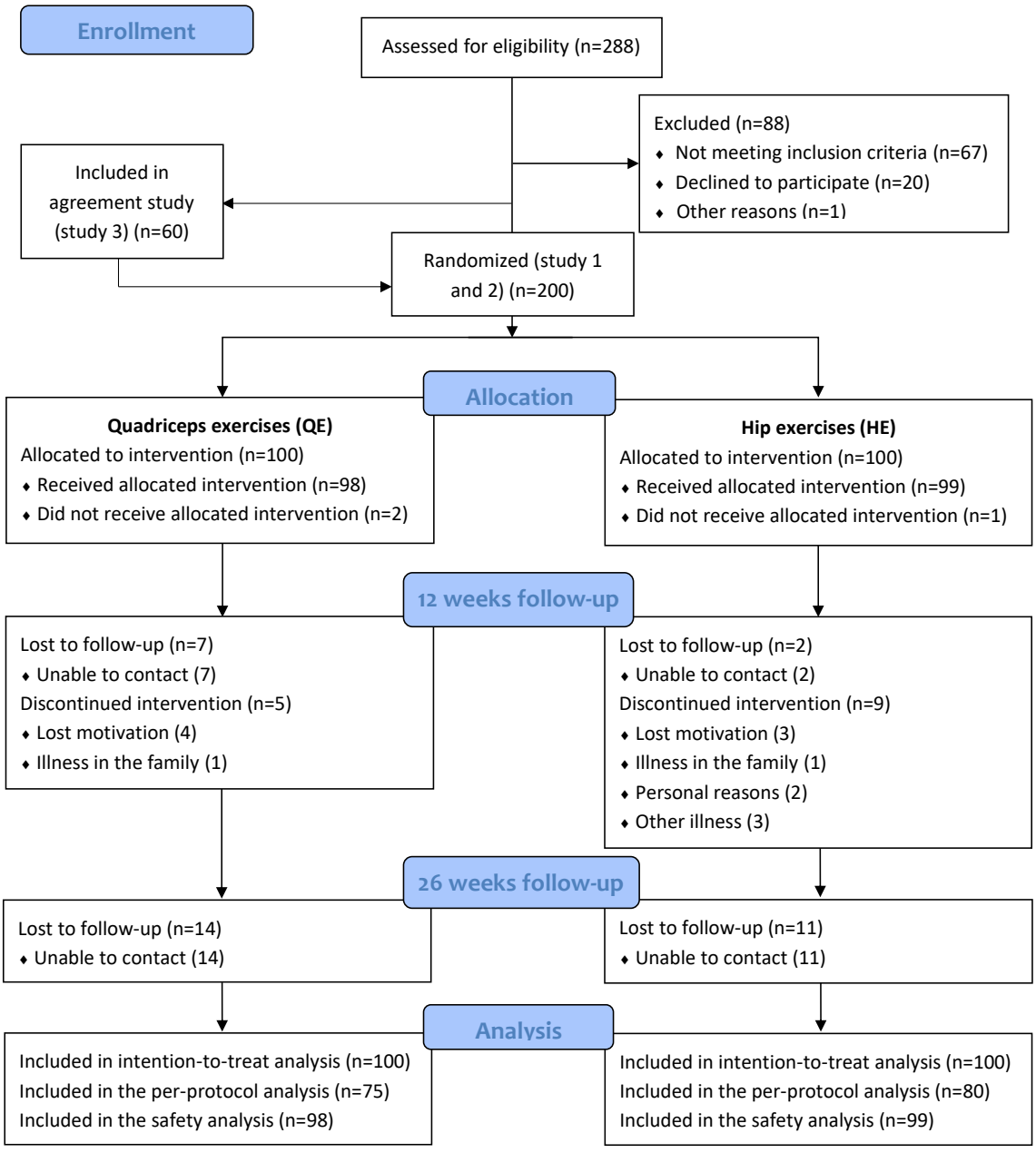
and clinical practice (113). The AKPS is a 13-item self-report questionnaire that documents response to 6 activities (walking, running, jumping, climbing stairs, squatting, and sitting for prolonged periods with knees bent), as well as symptoms such as limp, inability to weight bear, swelling, abnormal patellar movement, muscle atrophy, and limitations in knee flexion. The questionnaire ranges from 0 (delineating the worst function and symptoms) to 100 (full function and no symptoms) (117). The minimal clinical important difference (MCID) is reported to range from 8 to 19 points in patients with PFP (113, 116). The minimal MCID for VAS for usual pain is established at 1.5 to 2 cm and at 2 cm for the VAS for worst pain (113). Besides the AKPS, The Knee injury and Osteoarthritis Outcome Score (KOOS) is commonly used in PFP trials, but it was originally intended to be used for knee injury that can result in post-traumatic OA, i.e., anterior cruciate ligament injury, meniscus injury, chondral injury, etc. (119).

Different modalities of transition ratings or Global Rating of Change Scores (GRoC) are frequently used for the measure of improvements in a patient's condition. The participants are asked at follow-up to compare their current state with the state at baseline. This approach has limitations regarding recall bias and poor reliability over time, and fluctuation in results from week to week is a well-known downside (120).

Methods

The studies included in this thesis embrace different scientific designs and methods. Study 1 was the main study (an RCT) and Study 2 and 3 were sub studies. In the following, the design and methods of the individual studies are presented. A flowchart of the studies and included participants in the three studies is provided in figure 1.

Figure 1. Flowchart depicting the inclusion and exclusion rates with reasons for the three studies in the PhD thesis



Design

The RCT design is considered the gold standard for clinical trials and allows for valid conclusions on effect differences when conducted in accordance with acknowledged guidelines (121). In Study 1, we intended to compare two groups of patients performing two different exercise programs. The RCT design provides the most precise and valid data and was therefore chosen for the main study. Further, we used an equivalence framework as 1) our hypothesis was, that the two treatments (QE and HE) were equal in effectiveness, 2) it allowed for a more clinical applicable interpretation of the difference between treatments, and 3) it increased precision of the estimated differences. Equivalence is quantified by applying a tolerance margin often related to a clinically meaningful difference. If the mean difference and both ends of the 95% confidence interval of two groups fall within the tolerance range, the interventions are considered equivalent. Whereas a superiority trial is designed to detect a difference between treatments, the equivalence trial is designed to show that two treatments are not different in characteristics defined in a clinical manner (122). Further, it is possible to claim superiority with an equivalence design, but you can only rarely claim equivalence from a superiority design, primarily because of smaller sample sizes and wider confidence intervals (123). In PFP research, clinical trials are few and sample sizes often small, and therefore claims of superiority or no effect may not always be correctly interpreted. The equivalence design allowed us to draw reliable conclusions regarding the comparative effectiveness of the interventions and increases the precision of the estimated group differences.

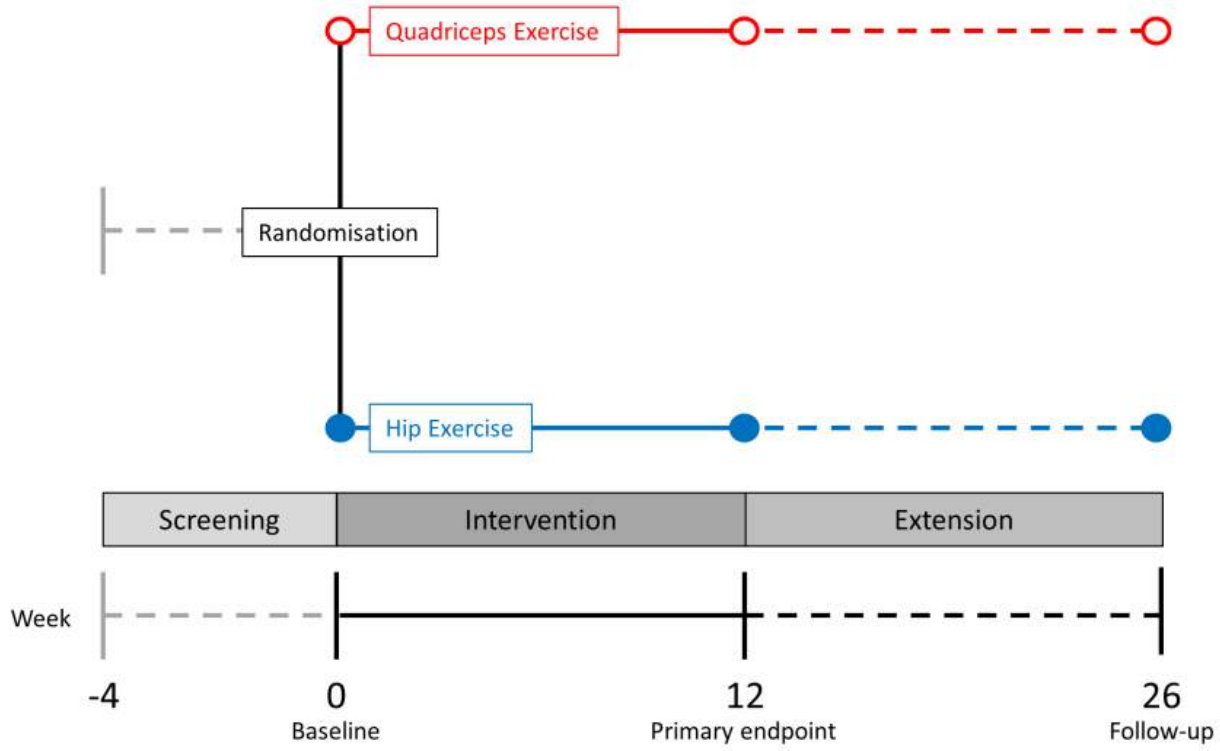
A description of the research designs used for each individual study is provided in the following:

Study 1

Study 1 was a single center, randomized, parallel-group, 26 weeks (6 months), equivalence trial comparing a 12-weeks focused QE protocol and a 12-weeks focused HE protocol with a primary endpoint at 12 weeks (after treatment) and a follow-up at 26 weeks. The trial was conducted in patients with PFP. The trial design is illustrated in figure 2.

An equivalence threshold for the primary outcome, the AKPS, were set to ±8 points, which is the established MCID for the PROM (113). That means that equivalence could be claimed if the observed difference in AKPS and its 95% CI fall inside the interval of clinical equivalence (±8 AKPS points).

Figure 2. An illustration of the trial design for the main study (Study 1)



Study 2

Randomized clinical trials provide evidence on the average effect of an intervention across the range of participants. However, it relies on the assumption that the effect of the intervention is consistent across all the participants in a trial. Stratified medicine refers to the division of a patient population into distinct subgroups based on particular characteristics. The PFP population is very heterogeneous and "one-size-fits-all"-approaches presumably are sub-optimal because the heterogeneity is ignored.

Study 2 was a secondary analysis of Study 1, exploring potential patient characteristics that predict differential responses to the two exercise programs (QE vs. HE) on self-reported pain and physical function. The analyses were based on assumptions that treatment effects vary according to an interaction between group allocation and certain baseline patient characteristics that can be defined dichotomously as present/absent. Continuous data was dichotomized when a reasonable and clinically applicable threshold could be set.

Study 3

This study was a cross-sectional study aiming to assess the inter- and intrarater agreement of visually assessing dynamic knee joint alignment during a single leg squat (SLS) and a forward lunge (FL). The population was a subset of the RCT study (Study 1) selected before

randomization. Dynamic knee alignment was included as a potential patient characteristic associated with treatment response (Study 2), and therefore we wanted to establish the consistency of the measurement process.

Reliability and agreement both refer to the issue of reproducibility of measures. Reliability assesses whether a test can differentiate between participants when measured twice under the same conditions, i.e., measurement variability. Agreement is the capacity of a measurement tool applied twice on the same population under the same conditions to provide strictly identical results, i.e., the measurement error in repeated measurements (124). Since therapists use visual assessments of lower extremity alignment in the clinic and in research, it is therefore relevant to examine the extent to which therapists can reproduce assessments and thereby trust their measures. It was out of the scope of this study to assess the reliability or the clinical validity as it would require repeated testing sessions and comparisons with advanced and expensive equipment.

Randomization technique, allocation concealment and blinding

Randomization technique describes the method of dividing the study sample into comparable groups, in order to balance the effect of confounders and to avoid systematic differences between treatment groups (125). Allocation concealment is defined as "*A technique used to prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. Allocation concealment prevents researchers from (unconsciously or otherwise) influencing which participants are assigned to a given intervention group*" (126).

Blinding or masking is essential for the internal validity of RCTs. It is intended to limit the occurrence of conscious or unconscious bias in the conduct and interpretation of a clinical trial. Potential biases can be the influence which the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of endpoints, the handling of withdrawals, the exclusion of data from analysis, and so on.

In the following, randomization, allocation concealment and blinding methods are described for the individual studies:

Study 1 and 2

In Study 1, the study director (MH) was responsible for preparation of the randomized group allocation list. Eligible participants were randomly assigned to one of the two groups (QE or HE)

in permuted blocks of 4 and 6 according to a computer-generated list of random numbers. The block randomization was chosen to ensure a balance in sample size across groups (125, 127, 128). The randomization was equal, meaning that 100 participants were randomly assigned to each group. Prior to the study initiation, a folder containing 200 envelopes was created and each envelope contained a piece of paper with the randomization written on it representing the group allocation (i.e., "QE", or "HE"). The order of the envelope content matched the randomization list. It was ensured that the envelopes were closed and opaque. The folders were stored in a locked locker in the principal investigator's (RH) office. Duplicates of the randomization list and envelopes were stored under lock in the Department of Physical and Occupational Therapy and delivered sequentially to the study physiotherapist at randomization. After inclusion of a participant, collection of signed informed consent from the participant, and completed baseline measurements, the clinical staff allocated participants according to the randomization list.

In the main RCT, the investigators, study coordinators, and the outcome assessor were blinded to allocation, and participants were requested not to disclose allocation during clinical assessments. Participants and staff involved in the exercise delivery were not blinded to the group allocation. Information that could potentially unblind otherwise blinded staff was not shared and was stored in facilities with limited access until the study was completed.

Study 3

The agreement study was a cross-sectional study, and the participants were sequentially asked to participate upon agreement to participation in the main RCT (Study 1). The first sixty individuals with PFP who were included in the RCT accepted the invitation to participate in the agreement study which minimizes the potential risk of selection bias.

The re-assessments (intra-rater assessment) were performed after a minimum of one week, so that any recall bias would be minimized. The mean time from baseline to re-evaluation was 29.1 days (SD 14.8). The visual analysis of the video recordings by the two raters (inter-rater assessment) was performed separately and without any interaction as recommended (129). Ratings from the two raters were saved on separate external hard drives and stored in facilities with limited access until the data collection was completed.

Participants and settings

The sample to be studied must be appropriate to the hypothesis being tested so that any results are appropriately generalizable (external validity). Patellofemoral pain patients are a heterogenous group consisting of both elite athletes, recreational exercisers, and sedentary

individuals (27), and hence it is important to set inclusion and exclusion criteria defining target populations that are appropriate to the research hypothesis.

Participants were recruited from the Institute of Sports Medicine Copenhagen (ISMC), Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark. Patients were eligible to participate in the study if they were diagnosed with PFP, if their symptom onset were insidious, and if the average pain during daily living exceeded 2 on a verbal 0-10 rating scale. Exclusion criteria included suffering from other knee disorders. A full description of in- and exclusion criteria is provided in the study protocol (Appendix 1)

The setting for the rehabilitation is another issue that concerns the generalization of the results. In the RCT, the setting resembled a typical rehabilitation in a Danish outpatient clinic, where patients most often do home based exercises and have weekly or monthly supervision visits at the clinic.

Sample size and power

Study 1 and 2

The sample size was calculated to test the equivalence of the QE and HE programs in the assessment of change in the AKPS questionnaire. With 77 participants per group, the study had 90% power and a significance level of 5% with the following assumptions: The expected group difference in mean changes from baseline was 0, the common standard deviation was 15 (0-100 scale), and the delta (equivalence margin) was 8 units (0-100 scale) corresponding to the suggested minimum clinically relevant difference. With an expected drop-out of 20% during the study we randomized and allocated 200 participants (100 to each group). We had no presumptions about the multiple group*predictor interactions that we pre-specified to explore in Study 2. However, we believed that the conservatively set sample size had sufficient power to reliably detect candidate characteristics that associate with a differential treatment response.

Study 3

In the agreement study we aimed to include 60 participants, which gave 80% power to detect a kappa-coefficient of at least 0.5 that is statistically significantly different from 0. This corresponds to a moderate agreement.

Interventions and procedures

Study 1 and 2

The two exercise interventions to be compared was a quadriceps focused and a hip-muscles focused resistance program. Both the hip and knee focused exercise programs were inspired by previous research (130) and followed recommended prescribing guidelines (131, 132). The duration of the exercise intervention was 12 weeks including three weekly exercise sessions consisting of three exercises (each with three sets of 8-12 repetitions). At least 48 hours of restitution was recommended between training sessions. The programs were home-based and transferable to a clinical setting in which it is not possible to monitor the patients on a daily or weekly basis. A monthly follow-up by a physiotherapist (clinical visit) with adjustment of technique and intensity was scheduled. The physiotherapists involved in instructing the patients (n=5) were all experienced in treating patients with PFP. They were instructed to communicate in the same vein, and training sessions were held in the planning stage to ensure a homogeneous communication and practice.

At the first clinical visit (after allocation), the participants were introduced to the exercises by the physiotherapist according to the allocation. The aim of the instruction was that the participant was able to do one set of each exercise with satisfying quality (i.e., full range of motion and without any compensatory movements throughout the entire set). The last repetitions should be difficult to perform while still allowing the participant to maintain high quality of movement. Participants were instructed to perform the exercises at a moderate velocity, i.e., 1-2 seconds in the concentric movement and 1-2 seconds in the eccentric movement. The load was set at 60-70% of 1 repetition maximum (RM), i.e., 8-12 repetitions. The recommended volume was 3 sets with 1½ - 2 minutes rest between sets (132). The use of elastic bands, free weights, and body weight as exercise resistance makes it impossible to estimate the exact RM. Therefore, participants were instructed to approximate fatigue within 8-12 repetitions in each set. An increase in resistance was recommended when the participant was able to perform 2 repetitions more than the desired number (i.e., 14 or more) (132) with satisfying quality. Key parameters of the two exercise programs are summarized in table 1. Detailed descriptions of the individual exercises of the two programs are summarized in Appendix 2.

Table 1. Key parameters of the exercise programs

	Number of repetitions/ sets	Time under tension	Rest in between sets	Means of progression	Exercise interventions per week
QE-1: Sitting leg extension	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands on ankles	3/week
QE-2: Squat	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding weight in a backpack (e.g. sand, flour, bottles of water) or by holding dumbbells in the hands.	3/week
QE-3: Lunge	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	As above	3/week
HE-1: Clam-shell	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands just above both knees	3/week
HE-2: Side-lying/standing hip abduction	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	As above	3/week
HE-3: Standing hip extension	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands from underneath the foot to the knee of the moving limb	3/week
QE: Quadriceps exercise HE: Hip exercise Both groups were instructed to warm up by performing 20 repetitions of exercise QE-1 (for QE) or HE-1 (for HE) without external load. Both groups were instructed to increase resistance whenever the participants were able to perform 2 repetitions more than the desired number (i.e., 14 or more)					

Information leaflet

All participants received the information leaflet "*Managing my patellofemoral pain*" (Danish title: "*Håndtering af mine forreste knæsmerter*"). The leaflet is targeted patients and written in an easy read language. It contained general information about causes and management of PFP. The leaflet is appended in Appendix 3 (Danish version).

Concomitant treatments

Other exercise programs/regimes than the one the participants were allocated to, were not allowed to be initiated during the main trial phase (week 1-12). Other non-pharmacological treatments were allowed. The usage of such other treatments/therapies was recorded. Habitual use of pharmacological therapies was allowed and was recorded. Any new pharmacological therapies or changes in ongoing therapies were recorded.

Post-intervention period

After the intervention period and after the 12 weeks outcome assessments, participants were advised to continue an active rehabilitation by continuing the prescribed exercises or by combining with other exercises, e.g., exercises from the other intervention group. However, adherence to exercises or any other rehabilitation was not monitored.

Adherence

Adherence to the prescribed exercise protocol was monitored by a self-administered exercise diary. The participants were asked to record date, number of repetitions and sets for each exercise, and the resistance (i.e., elastic band color corresponding to a specified resistance or weights in kilograms) for each exercise session. Adherence was assessed based on the percent of the scheduled number of training sessions that was performed. The number of scheduled training sessions for both intervention groups was predefined in the trial protocol and equaled 36 sessions for 12 weeks. A training session was considered performed if an exercise activity was registered at a given date, even if the repetitions, sets, or exercises were only partly performed. The pre-defined threshold for satisfactory intervention adherence was 24 out of 36 scheduled training sessions (66%).

Study 3

Visual observation of functional tests is the most commonly used method of assessing dynamic alignment in the clinic (69). Furthermore, functional tests are a simple way of assessing movement quality through visual observations (65). Clinicians use visual ratings to make clinical decisions and to evaluate progress of rehabilitation. Therefore, the measurement properties of these assessments need to be considered. The assessment method used in this study resembles a clinical setting, where clinicians have limited access to the equipment and limited time for complex biomechanical analyses. The SLS and the FL were chosen because they are commonly used in clinical practice and have been reported in many previous studies investigating visual rating of lower extremity function (69, 133). The raters in this agreement study were two sports physiotherapists with 18 and 15 years of experience, respectively, in treating and assessing patients with musculoskeletal problems.

General procedures

Video was recorded using a tablet from an anterior view of participants performing the SLS and FL in the gym at the Department of Physical and Occupational Therapy at Bispebjerg and Frederiksberg Hospital. An investigator instructed the participants to perform the test as described below. The participant performed the selected movement with no prior rehearsal. If the participants lost their balance during the test, a new attempt was initiated. The participants remained unaware of the classification criteria during testing. The investigator video recorded the participant and scored the movement according to the criteria set below and in table 2. The video captured the whole person. At least 1 week later, the investigator did another scoring based

on the recorded video and another investigator repeated the scoring. Three playbacks of the recorded video in real time were allowed for the intra- and interrater assessment.

Knee alignment during single leg squat and forward lunge

The SLS test was performed as follows: The participant stood on one leg (painful knee) and performed a squat (hip and knee flexion) while maintaining the trunk in an upright position, the contralateral hip in neutral, and the contralateral knee flexed. The participant was instructed to perform the squat until reaching maximal ankle dorsiflexion without lifting the heel and return to the upright starting position (figure 3a). In the FL test, the participant stood on both feet and performed a forward step (on painful knee). The participant then flexed the knees simultaneously (forward lunge). The participant was instructed to continue the lunge until reaching maximum dorsiflexion of the front ankle without lifting the heel and then push-off to upright starting position (Figure 3b). Both tests were performed at a participant-selected speed. Dynamic valgus alignment was defined as an excessive medial movement of the knee as evidenced by an increased frontal plane knee angle during the selected movement. Varus alignment was defined as an excessive lateral movement. 'Severe' valgus/varus was defined as a clinically relevant and severe collapse of the knee, 'moderate' valgus/varus was defined as a modest and clinically relevant deviation from neutral, 'mild' valgus/varus was defined as a slight deviation from neutral that might not be clinically relevant, and 'doubtful' was defined as a just merely detectable deviation from neutral alignment. We defined `no evidence of neither valgus nor varus' as a neutral knee alignment, i.e., knee flexion aligned with the 2nd toe. The scores and classifications are summarized in table 2.

Figure 3. Screenshots of video recording capturing a model performing an (a) single-leg squat and (b) a forward lunge

Table 2. Scores and classifications

Score	Definition and Classification	
-4	Severe valgus	
-3	Moderate valgus	Definite valgus present
-2	Mild valgus	
-1	Doubtful valgus	
0	No evidence of either valgus or varus	No evidence of dynamic malalignment
1	Doubtful varus	
2	Mild varus	
3	Moderate varus	Definite varus present
4	Severe varus	

General ethical considerations

The protocol and written patient information was submitted to the ethical committee, prior to study initiation. Ethics approval was given by the Health Research Ethics Committee of the Capital Region of Denmark (H-16045755). The three studies were conducted in accordance with Danish law, the Helsinki declaration, and local research ethics committee requirements.
Prior to screening, all potential trial participants were informed, both orally and in writing, about the trial's purpose, process and potential risks, costs and benefits of participation. Participants written informed consent were obtained prior to the start of the studies.

Outcomes

The outcomes for each individual study are summarized in the following section. An overview of the outcomes including the timing is provided in table 3 and the full description and rationale for each outcome measure is provided in the study protocol (Appendix 1)

Table 3. The outcome measures and timing for the three studies

Study	Baseline	Week 12	Week 26
1	• AKPS • KOOS • Dynamic Assessment of Pain • Pain Self-efficacy Questionnaire • EuroQoL EQ-5D-3L • Isometric muscle strength	• AKPS (primary) • KOOS • Dynamic Assessment of Pain • Pain Self-efficacy Questionnaire • EuroQoL EQ-5D-3L • Isometric muscle strength • GRoC	• AKPS • KOOS • Pain Self-efficacy Questionnaire • EuroQoL EQ-5D-3L • GRoC
2	• AKPS • Baseline characteristics, e.g. pain severity, BMI, Pain duration etc.	• AKPS	• AKPS
3	Intra- and interrater agreement when visually assessing the Single Leg Squat and the Forward Lunge (cross sectional)		

Study 1

Primary outcome

The primary outcome for Study 1 was assessed at week 12 as change from baseline in the AKPS questionnaire (117). We analyzed the group difference in the mean changes from baseline in the AKPS questionnaire between QE vs HE after 12 weeks.

Key Secondary outcomes

The following outcomes were assessed as key secondary outcomes:

- Change from baseline in the KOOS (119) pain subscore at week 12
- Change from baseline in the KOOS function subscore at week 12
- Change from baseline in the KOOS quality of life subscore at week 12

Other secondary outcomes

The following outcomes were assessed as other secondary outcomes:

- Change from baseline in the KOOS Symptoms and Sports/Recreation subscores at week 12
- Change from baseline in isometric muscle strength of hip abductors, hip adductors, hip external rotators, hip internal rotators, hip extensors, hip flexors, knee flexors (hamstrings), and knee extensors (quadriceps) (134-136) at week 12
- Change from baseline in Dynamic Assessment of Pain (137) at week 12
- Change from baseline in Pain Self-Efficacy Questionnaire (138) at week 12
- Change from baseline in the 3-level version of EuroQoL 5 dimensions (EQ-5D-3L) Questionnaire (139) at week 12
- The Transition Questionnaire of global perceived effect on overall health, pain, and function measured on a 15-point Likert scale ranging from -7 (much worse) to +7 (much better) at week 12
- Change from baseline in the outcomes measured at week 26 (only questionnaire data)

Study 2

As Study 2 is a secondary analysis of Study 1, the primary outcome is change from baseline in the AKPS questionnaire at week 12 and week 26. A priori we identified a range of baseline variables to be explored as potential effect modifiers. We analyzed the interaction between group allocation (QE and HE) and the patient characteristics (binary defined as present/absent). The potential effect modifiers are presented in the following.

Potential effect modifiers

The candidate baseline characteristics were self-reported information as well as clinical observations and tests and were chosen based on findings in previous studies and clinical experience (20, 107): Presence of low back, hip, ankle, or bilateral knee pain, body mass index (BMI), sex, age, education, occupation, hypermobility, quadriceps strength, dynamic knee alignment, midfoot mobility, exercise self-efficacy, pain self-efficacy, pain catastrophizing, neuropathic pain, pain duration, and pain severity. The characteristics were dichotomized in accordance with established clinically relevant cut-off values where possible. For quadriceps strength and age, no meaningful cut-off values were found, and dichotomization was hence based on median values. The baseline characteristics are presented in table 4 and a full description of the characteristics including rationales for selecting the specific items and cut-off values is available in table A in the Statistical Analysis Plan (SAP) Appendix 5.

Table 4. Candidate baseline characteristics

Variables	Description
Low back pain	Self-reported presence of low back pain during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", "Monthly" were defined as "Low back pain present" and the scores "No" and "Rarely" as "Low back pain not present".
Presence of bilateral knee pain	Self-reported presence of pain in the contralateral knee (not target knee) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", "Monthly" were defined as "Pain in the contralateral knee present" and the scores "No" and "Rarely" as "Pain in the contralateral knee not present".
Presence of ankle pain	Self-reported presence of ankle pain (one or both ankles) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", and "Monthly" were defined as "Ankle pain present" and the scores "No" and "Rarely" as "Ankle pain not present".
Presence of hip pain	Self-reported presence of hip pain (one or both hips) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", and "Monthly" were defined as "Hip pain present" and the scores "No" and "Rarely" as "Hip pain not present".
Body mass index (BMI)	The participants' BMI is measured at baseline and dichotomized with a cut-off at 25 (BMI $\geq$ 25).
Sex	Female vs. male sex
Duration of knee pain	Self-reported duration of knee pain (present condition). We defined chronic pain as pain lasting for 6 months or longer.
Education level	Self-reported highest level of education: Scores were dichotomized: "Medium-term higher education (3-4 years)", and "Longer higher education (>4 years)" were defined as "Long education" and "Primary school", "Craftsman", "Highscool", and "Short higher education (<3 years)" as "Short education".
Occupation	Self-reported status of occupation and education. Scores were dichotomized: "Currently studying" and "Currently working" were defined as "Currently studying/working" and the score "Currently not working" as "Currently not working/not studying".
Hypermobility	Hypermobility assessed by the Beighton Score applying the revised criteria for the diagnosis of benign joint hypermobility syndrome. The Beighton score ranges from 0-9. Scores were dichotomized: scores 4-9 were defined as "Generalized joint hypermobility" and scores 0-3 as "Normal joint mobility". A score of 4 or more, is generally considered an indication of joint hypermobility (140, 141).
Quadriceps strength	Quadriceps strength is measured by handheld dynamometry. The muscle strength tests are conducted following validated testing protocols (134-136). As there is no established threshold available for sufficient or adequate muscle strength in PFP patients, measures were dichotomized according to the median strength (274 N) of all trial participants in the ITT population
Knee joint valgus malalignment during a forward lunge movement	Knee joint valgus malalignment assessed by clinical observation of the participant while he/she performs a forward lunge movement. The scores were dichotomized: The score "Definite valgus present" was defined as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".

Variables	Description
Knee joint valgus malalignment during a single-leg squat movement	Knee joint valgus malalignment assessed by clinically observation of the participant while he/she performs a single-leg squat movement. The scores were dichotomized: The score "Definite valgus present" was defined as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".
Exercise self-efficacy	Self-reported exercise self-efficacy in relation to the two different exercise programs assessed by asking the participants to rate their confidence in performing the allocated exercise program on an 11-point (0-10) Likert scale with 0 representing "Not at all confident" and 10 representing "Completely confident". Scores were dichotomized: scores 6-10 were defined as "High self-efficacy" and the scores 0-5 as "Low self-efficacy".
Neuropatic pain	Presence of signs of neuropathic pain assessed by The painDETECT questionnaire (PDQ). A validated algorithm is used to calculate a total score ranging from -1 to 38. Scores were dichotomized: Scores ≥19 were defined as "Neuropathic pain component" and scores ≤18 as "No neuropathic pain component" as recommended by Freynhagen et al. (142).
Pain severity	Self-reported average pain during the past 4 weeks on a 0-10 Numeric Rating Scale. Scores were dichotomized: scores 0-6 were defined as "Mild or moderate pain" and the scores 7-10 as "Severe pain".
Pain Catastrophizing Scale	Presence of pain-related catastrophic thinking assessed by The Pain Catastrophizing Scale (PCS). Scores were dichotomized: scores > 30 were defined as "Pain catastrophizing" and scores 0 - 30 as "No pain catastrophizing". Previous studies have shown a cut-off of more than 30 points to be of clinical relevance (143, 144).
Pain self-efficacy	Confidence in performing activities while in pain assessed by The pain self-efficacy questionnaire. Confidence in performing activities is rated on a 7-point (0-6) Likert scale with 0 representing not at all confident and 6 representing completely confident. A total score is calculated by summing the answers producing a score between 0 and 60. Scores were dichotomized: scores 0-39 was defined as "Poor pain self-efficacy" and the scores 40-60 as "Good self-efficacy". Scores around 40 (percentile = 50) are associated with return to work and maintenance of functional gains, whilst lower scores tend to predict less sustainable gains in injured workers (145).
Midfoot mobility magnitude	Midfoot mobility measured by the change in midfoot width from non-weight bearing to weight bearing. We defined a dichotomization of scores within the range 0-1.24 cm (0.92 cm +1*SD) as "Normal or limited midfoot mobility magnitude" and the scores above 1.25 cm as "Excessive midfoot mobility magnitude" for the females. For the males, we defined a dichotomization of scores within the range 0-1.36 cm (1.02 cm +1*SD) as "Normal or limited midfoot mobility magnitude and the scores above 1.37 cm as "Excessive midfoot mobility magnitude". Dichotomization is based on normative data for foot mobility (146).
Age	Age at inclusion in the trial. To split the trial population in two groups, we chose to categorize the participants based on the median age (26 years) of all trial participants in the ITT population.

Study 3

In this cross-sectional study we aimed to assess the intra- and intertester agreement of two commonly used assessments (SLS and FL) for the evaluation of dynamic knee alignment.

Statistics

Study 1

The statistical analyses were all pre-specified in the SAP (Appendix 4). All participants were analyzed according to the intention-to-treat (ITT) principle. We used the repeated measures linear mixed model regression analysis model adjusted for the baseline score of the AKPS questionnaire. *Equivalence* was claimed if the computed 95% confidence interval of the estimated group difference in the change from baseline in the AKPS questionnaire at week 12 did not include ± 8 AKPS points in the primary analysis. *Superiority* was claimed if the computed 95% confidence interval of the estimated group difference in the change from baseline in the AKPS at week 12 did not include 0 in the primary analysis. We performed secondary sensitivity analyses to test the robustness of the results (SAP, Appendix 4).

Study 2

The statistical analyses were all pre-specified in the SAP (Appendix 5). In the primary analysis, all participants were analyzed using the ITT population according to the intention-to-treat principle. The analyses focused on the change from baseline in the AKPS at week 12 and 26. We used a repeated measures linear mixed model regression analysis model adjusted for the baseline score of the AKPS. An interaction for time, week, "modifier covariate" and group was included. Any group difference in the change from baseline in AKPS between sub-groups of participants based on the presence of the potential effect modifiers were estimated together with the associated 95% confidence interval. All 95% confidence intervals and *P*-values were two sided. We did not apply explicit adjustments for multiplicity, rather we explicitly state that the results are exploratory and hypothesis generating.

Study 3

For this study we used Cohen's weighted kappa statistics to assess the intra- and interrater agreement. Cohen's kappa is a robust statistical method for interrater or intrarater agreement testing when the type of variable is categorial (147). It ranges from −1 to +1, where 0 represents the amount of agreement that can be expected from random chance, and 1 represents perfect agreement between the raters. For interpretation, we used the definitions adapted from Landis et al. (148), where kappa values below 0.20 signifies poor agreement; 0.21-0.40 signifies fair agreement; 0.41-0.60 signifies moderate agreement; 0.61-0.80 signifies good agreement; and 0.81-1.00 signifies very good agreement. Agreement was assessed using the classification for each movement test, i.e., valgus, no malalignment, or varus, and for raw data (scores from -4 to 4).

Summary of findings

This chapter sets out the findings of each study.

Study 1

In total, 288 individuals were screened for eligibility; 88 were ineligible for inclusion. Thus, 200 subjects underwent randomization; 100 were assigned to QE and 100 to HE. The mean age was 27.2 years; 69% were females; and the mean BMI was 22.6. Baseline characteristics were similar in the two groups (table 5).

Table 5. Baseline characteristics

	Quadriceps exercise group (QE) N=100	Hip exercise group (HE) N=100
Demographics		
Age, years	27.2 (6.3)	27.2 (6.7)
Female sex (n[%])	66 (66%)	72 (72%)
Body mass, kg	68.2 (12.4)	67.6 (13.0)
Height, cm	172.4 (8.5)	173.2 (10.7)
Body Mass Index, BMI (kg/m2)	22.8 (3.01)	22.4 (2.9)
Symptom duration, months	47.3 (49.4)	52.8 (54.1)
Symptom duration, months (median (IQR))*	36 (48)	30 (60)
AKPS questionnaire score (0-100)	74.2 (11.6)	73.3 (13.0)
KOOS (0-100)		
Pain	70.8 (15.6)	72.2 (14.1)
Physical Function	84.1 (13.2)	83.4 (13.1)
Symptoms	79.6 (14.0)	80.7 (13.3)
QoL	44.4 (15.1)	44.2 (14.7)
Sports & Recreation	56.7 (24.9)	59.3 (24.4)
Dynamic assessment of pain (VRS 0-10)	1.9 (2.2)	1.8 (1.8)
Dynamic assessment of pain (VRS 0-10) (median (IQR))*	1.0 (3.5)	2.0 (3.0)
Isometric muscle strength		
Hip abductors (N)	129.5 (40.9)	129.5 (41.2)
Hip adductors (N)	121.2 (40.6)	122.4 (47.3)
Hip extensors (N)	175.3 (46.8)	181.0 (56.0)
Hip flexors (N)	189.1 (55.8)	194.3 (63.0)
Hip external rotators (N)	101.1 (30.2)	100.7 (37.7)
Hip internal rotators (N)	123.9 (100.9)	109.4 (42.8)
Knee extensors (quadriceps) (N)	299.2 (113.1)	292.0 (121.9)
Knee flexors (hamstrings) (N)	316.6 (117.1)	302.8 (129.5)
Pain Self-efficacy questionnaire (0-60)	47.5 (8.6)	46.8 (9.8)
EuroQoL EQ5D Questionnaire (-0.624 to 1.000)	0.755 (0.175)	0.757 (0.127)

Values are presented as means and standard deviations (SD) unless otherwise stated.
*** Both means (SD) and medians (IQR) are presented as data is not normally distributed.**
IQR: Inter quartile range
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

For the primary outcome, the mean changes in AKPS questionnaire score from baseline to week 12 were 7.5 (SE ±0.8) in the QE group and 7.2 (SE ±0.8) in the HE group (table 6). The 95% CI of the group difference in change in AKPS questionnaire from baseline to week 12 was within the predefined equivalence margin of ±8 points; p<0.0001 for equivalence. Group difference was 0.3 points, 95% CI −1.9 to 2.4; p=0.804 for test of superiority. The trajectories of the AKPS questionnaire are shown in figure 4.

For the secondary outcomes, the treatment differences between groups at week 12 were 3.0 points (95% CI 0.1 to 5.9) for KOOS pain score, 0.6 points (95% CI −1.7 to 3.0) for KOOS function, and -1.5 points (95% CI −5.4 to 2.5) for KOOS quality of life score (positive treatment differences deflects an outcome in favor of QE, whereas a negative number is in favor of HE). The key secondary outcomes were all within the predefined criteria for equivalence, although the between group difference for KOOS pain was statistically significant in favor of QE (table 6). Finally, the results in the primary and key secondary outcomes appeared unchanged at week 26 (Table 7). The overall pattern of results for all outcomes was unchanged in the sensitivity analyses (tables S1-S3, Appendix 6).

Adverse events were mostly mild to moderate, mostly related to muscle soreness, and were similar in the two groups (Appendix 7). Severe adverse, defined as events that gave interference with the participants' usual activities, were also equally distributed in the intervention groups.

Figure 4. Trajectories for the AKPS

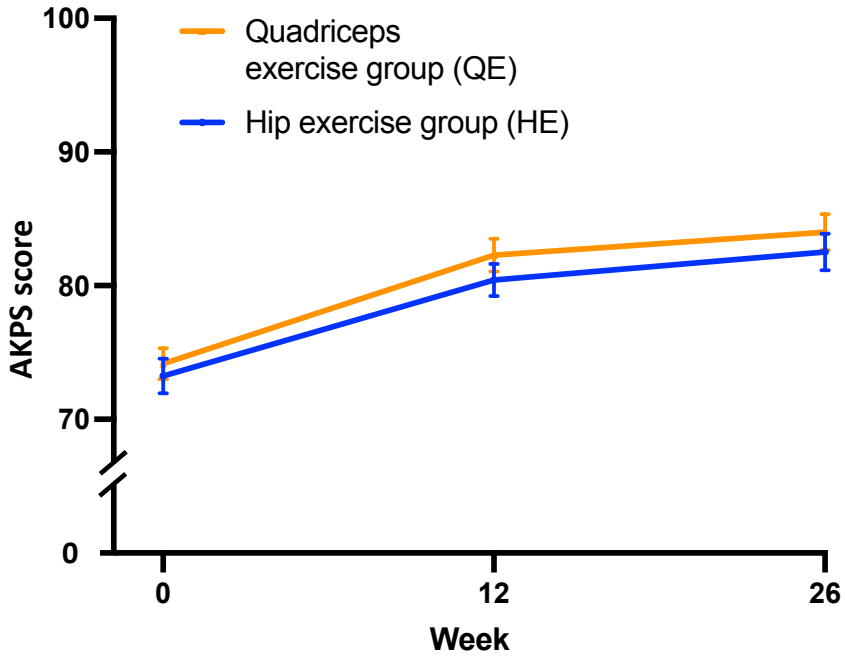


Table 6. Primary and Secondary Outcomes at week 12 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (95% CI)	P-value
Primary outcome:				
Change in AKPS questionnaire score (0 to 100); equivalence test*	7.5 (0.8)	7.2 (0.8)	0.3 (-1.9 to 2.4)	<0.0001
Change in AKPS questionnaire score (0 to 100); superiority test*				0.804
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	9.4 (1.0)	6.4 (1.0)	3.0 (0.1 to 5.9)	
Change in KOOS Function – score (0 to 100)	5.7 (0.9)	5.1 (0.9)	0.6 (-1.7 to 3.0)	
Change in KOOS Quality of life – score (0 to 100)	10.7 (1.4)	12.2 (1.4)	-1.5 (-5.4 to 2.5)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	13.8 (1.7)	11.0 (1.7)	2.8 (-1.9 to 7.6)	
Change in KOOS Symptoms – score (0 to 100)	4.8 (0.8)	4.9 (0.8)	-0.1 (-2.3 to 2.1)	
Change in isometric muscle strength				
Hip abductors (N)	13.7 (1.8)	13.3 (1.9)	0.4 (-4.7 to 5.5)	
Hip adductors (N)	10.7 (1.9)	16.3 (1.9)	-5.5 (-10.8 to -0.3)	
Hip extensors (N)	16.4 (2.5)	13.7 (2.5)	2.6 (-4.4 to 9.6)	
Hip flexors (N)	11.8 (2.0)	11.2 (2.0)	0.6 (-5.0 to 6.2)	
Hip external rotators (N)	1.7 (4.3)	8.4 (4.4)	-6.7 (-18.8 to 5.4)	
Hip internal rotators (N)	9.4 (1.5)	10.6 (1.5)	-1.3 (-5.5 to 3.0)	
Knee extensors (quadriceps) (N)	33.3 (5.6)	33.3 (5.7)	-0.1 (-15.8 to 15.7)	
Knee flexors (hamstrings) (N)	37.6 (4.1)	42.1 (4.1)	-4.4 (-15.8 to 6.9)	
Change in Dynamic Assessment of Pain (VRS (0-10))	-0.8 (0.1)	-0.2 (0.1)	-0.6 (-0.9 to -0.3)	
Change in EQ5D Questionnaire (index -0.624 to 1.000)	0.067 (0.011)	0.035 (0.011)	0.024 (-0.009 to 0.057)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.1 (0.3)	2.1 (0.3)	0.0 (-0.8 to 0.8)	
Treatment adherence				
Treatment adherence (%)	75.0 (23.2)	79.0 (21.3)	-4.0 (-10.2 to 2.2)	
Treatment adherers (adherence ≥66%) - no. (%)	82 (82.0%)	85 (85.0%)		

*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

Table 7. Primary and Secondary Outcomes at week 26 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (CI)	P-value
Primary outcome:				
Change in AKPS questionnaire – score (0 to 100); equivalence test*	9.8 (0.8)	9.0 (0.8)	0.9 (-1.4 to 3.1)	<0.0001
Change in AKPS questionnaire – score (0 to 100); superiority test*				0.449
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	10.7 (1.1)	10.4 (1.1)	0.4 (-2.7 to 3.4)	
Change in KOOS Function – score (0 to 100)	6.6 (0.9)	6.9 (0.9)	-0.3 (-2.8 to 2.1)	
Change in KOOS Quality of life – score (0 to 100)	15.9 (1.5)	19.0 (1.5)	-3.2 (-7.2 to 0.9)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	15.1 (1.8)	14.6 (1.8)	0.5 (-4.4 to 5.4)	
Change in KOOS Symptoms – score (0 to 100)	5.9 (0.8)	6.2 (0.8)	-0.3 (-2.6 to 2.0)	
Change in EQ5D Questionnaire (index -0.624 to 1.000)	0.093 (0.012)	0.069 (0.012)	0.024 (-0.009 to 0.057)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.6 (0.3)	2.7 (0.3)	-0.1 (-0.9 to 0.7)	

Group values for QE and HE are presented as least squares means ± standard error.
Mean differences are presented as least squares means and 95% confidence intervals (CI).
***Primary outcome was analyzed using both a test for equivalence and a test for superiority.**

Study 2

As this was a secondary analysis of Study 1, the baseline characteristics are the same and presented in table 5

The overall mean changes in AKPS score from baseline to week 12 and 26 are identical to the main trial results. At the 12-weeks follow-up, the group mean changes were 7.5 (SE ±0.8) in the QE group and 7.2 (SE ±0.8) in the HE group and at 26-week follow-up, the group mean changes were 9.0 (SE 0.8) and 9.8 (SE 0.8) in the QE and HE groups, respectively.

We found a statistically significant subgroup difference at 12 weeks in favor of QE among participants with the baseline characteristic “Pain catastrophizing” (n=22) compared to those without signs of pain catastrophizing at baseline (n=178) with a subgroup difference of 8.3

AKPS points (95% CI 1.6 to 15.0; p=0.016). At 26 weeks, participants with a baseline BMI ≥25 m/kg2 (n=32) seemed to benefit from QE compared to those with BMI <25 m/kg2 (n=168) with a subgroup difference of 11.1 AKPS points (95% CI 4.8 to 17.4; p=0.001). In contrast, participants with severe baseline knee pain (NRS >6) (n=28) seemed to benefit from HE when compared to those with mild-moderate baseline knee pain (n=172) with a subgroup difference of -9.1 AKPS points (95% CI -15.7 to -2.6; p=0.006). The results of the subgroup analyses are shown in figure 5 for the week 12 and figure 6 for the 26-week follow-up assessment.

Figure 5. Forest plots of the Treatment effect (QE vs HE) across different subgroups at week 12 based on dichotomous baseline variables (effect modifiers) in the ITT population.

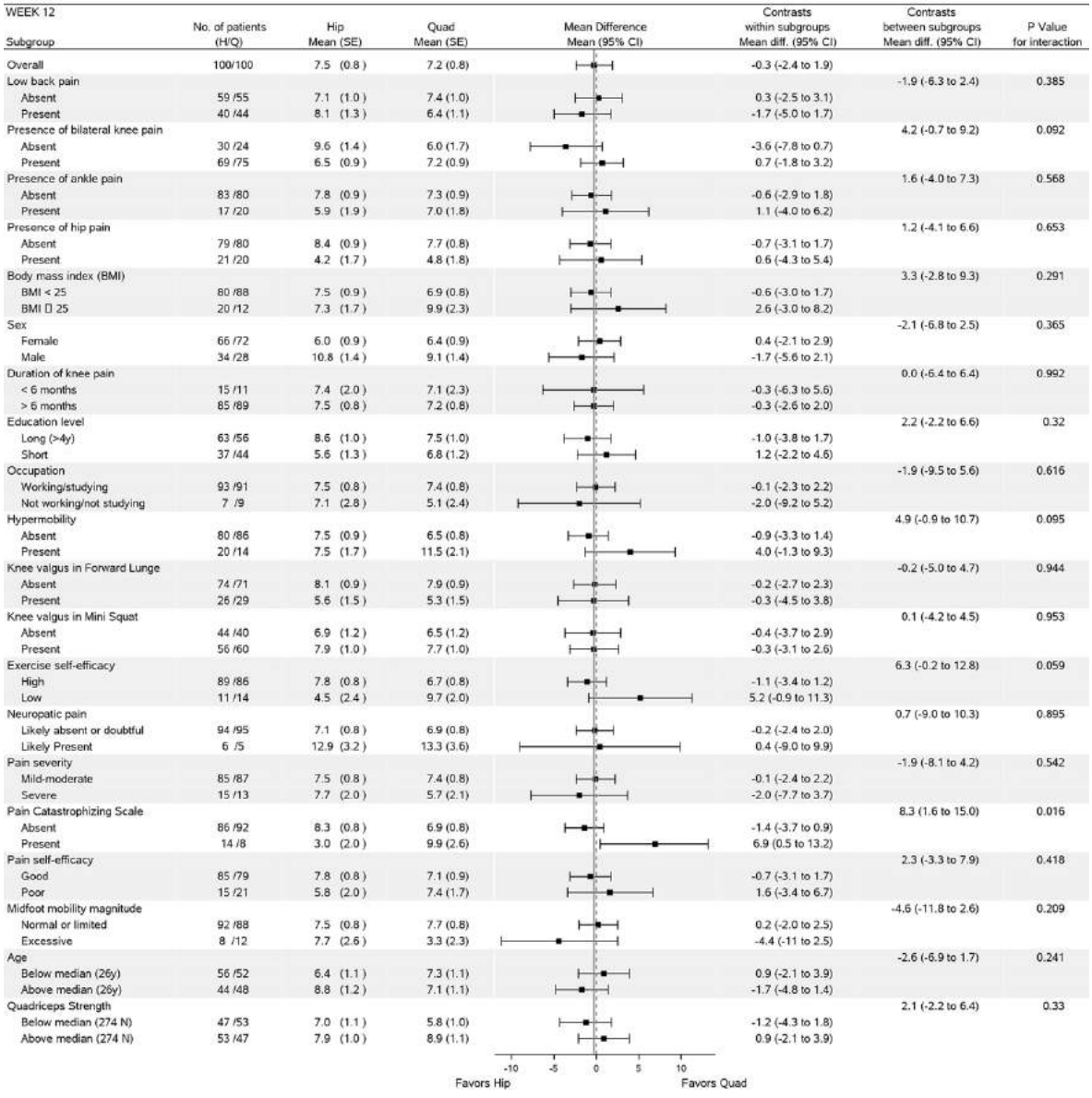
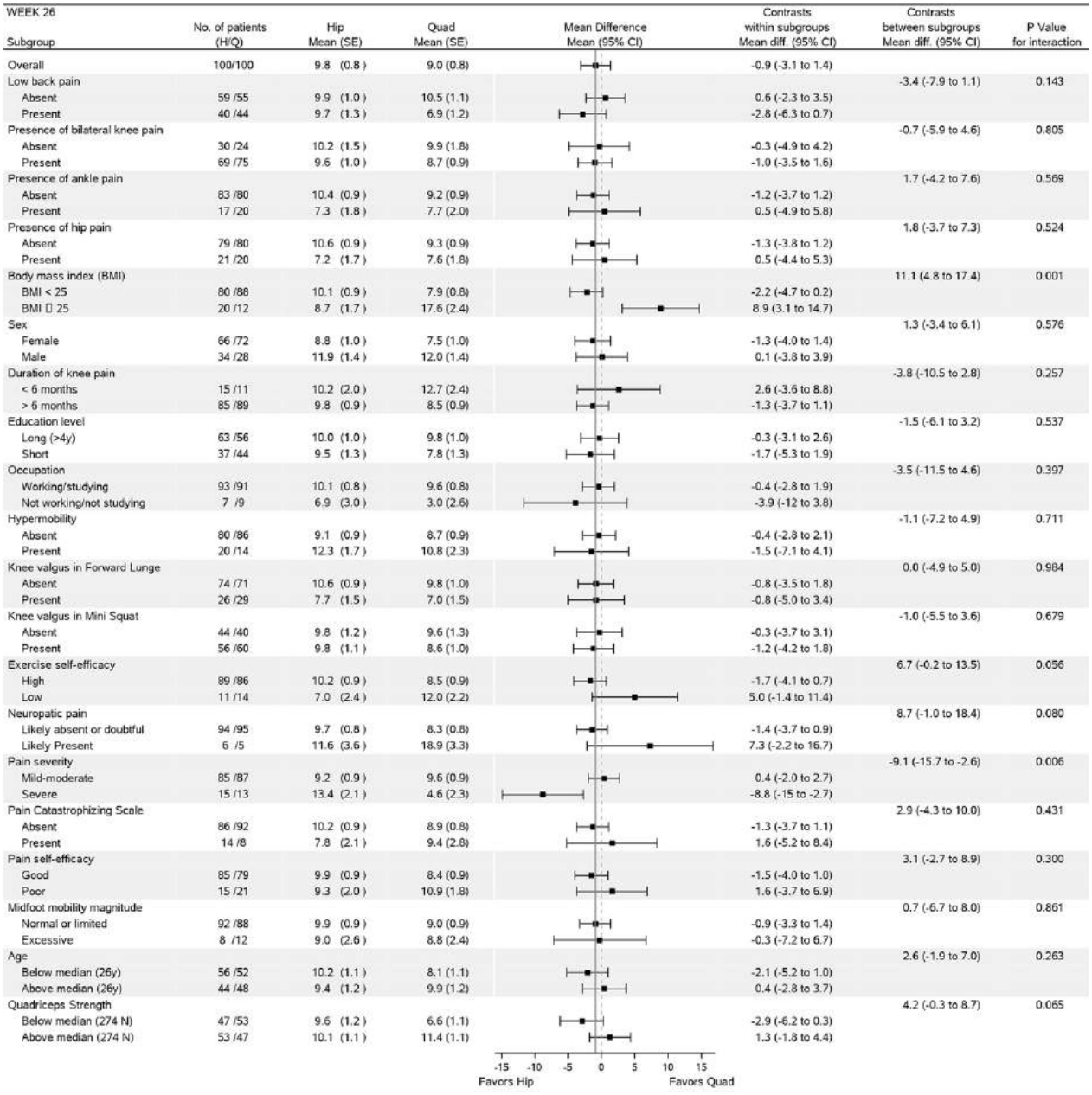


Figure 6. Forest plots of the Treatment effect (QE vs HE) across different subgroups at week 26 based on dichotomous baseline variables (effect modifiers) in the ITT population.



Study 3

The first 60 individuals with PFP who were included in the parent trial (Study 1) accepted the invitation to participate in this agreement study. Their characteristics are shown in Table 8. The dispersion of the data from the intra- and interrater assessments is presented in figure 7 (SLS) and figure 8 (FL). The weighted kappa values for the classifications and the raw scores are shown in tables 9 and 10. In summary, the intrarater agreement were statistically significantly different from 0 ($p<0.0001$) and ranged from 0.58 to 0.70, i.e., moderate to good agreement,

whereas the interrater agreement ranged from 0.22 (p=0.08) to 0.50 (p<0.0001), i.e., fair to moderate agreement. Interrater agreement was generally not as good as intrarater agreement (0.7 for SLS intrarater classification scores vs. 0.22 for interrater scores, and 0.58 for FL intrarater classification scores vs. 0.48 for interrater scores). The mean time from baseline to re-evaluation in the intrarater assessment was 29.1 days (SD 14.8). The cross tabulated agreements in the classifications and raw scores are provided in Appendix 8.

Table 8. Descriptive characteristics of the participants in Study 3

Characteristics	Mean (SD) N=60
Age (yrs)	27.2 (6.2)
Females (n (%))	42 (70%)
Height (cm)	172.1 (8.6)
Weight (kg)	66.6 (10.6)
Body mass index	22.4 (2.8)
Duration of symptoms (months)	34 (34)
Average pain during previous 4 weeks (NRS* 0-10)	3.73 (2.17)
*Numeric Rating Scale	

Figure 7. Heatmap of agreement matrix showing the dispersion of the intrarater (left) and interrater agreement (right) for the single leg squat. The brightness of the blue color indicates the number of rating combinations with darker colors representing higher numbers as shown in the individual squares and in the key bar.

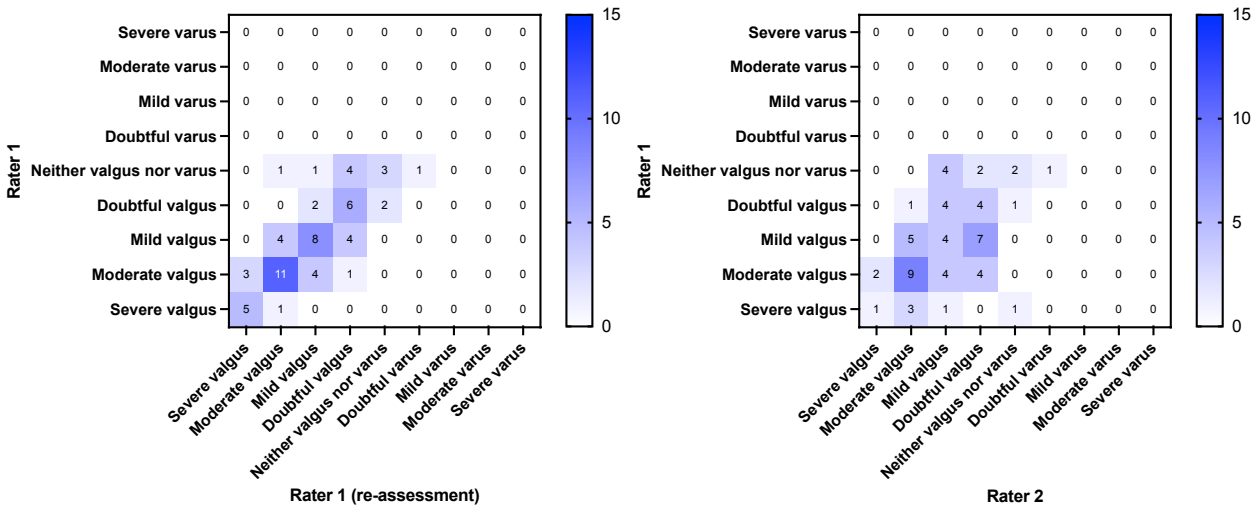


Figure 8. Heatmap of agreement matrix showing the dispersion of the intrarater (left) and interrater agreement (right) for the forward lunge. The brightness of the blue color indicates the number of rating combinations with darker colors representing higher numbers as shown in the individual squares and in the key bar.

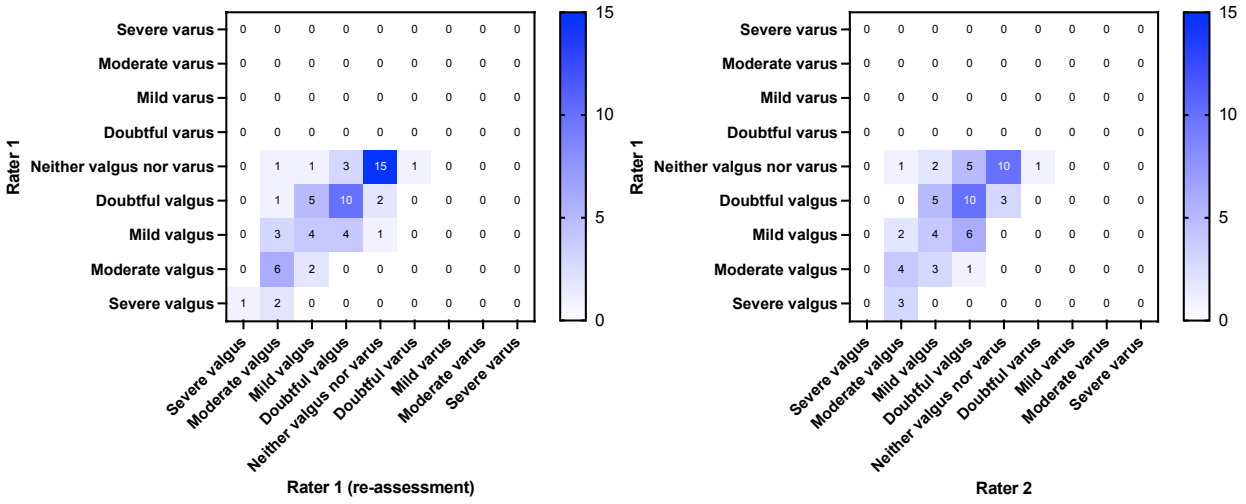


Table 9. Intra- and interrater agreement of single leg squat.

	Weighted kappa value	95% CI	p-value
Intrarater classification	0.70	0.51-0.89	<.0001
Intrarater raw data	0.65	0.54-0.76	<.0001
Interrater classification	0.22	-0.03-0.48	0.08
Interrater raw data	0.32	0.16-0.48	0.05

Table 10. Intra- and interrater agreement of forward lunge.

	Weighted kappa value	95% CI	p-value
Intrarater classification	0.58	0.37-0.79	<.0001
Intrarater raw data	0.65	0.53-0.78	<.0001
Interrater classification	0.48	0.25-0.7	0.0002
Interrater raw data	0.50	0.36-0.64	<.0001

General discussion

In Study 1, we found that exercise programs that focus on either quadriceps or hip muscles provide equivalent improvements in symptoms and function in the short (12 weeks) and medium term (26 weeks) in a population of PFP patients. Both exercise programs induced improvements in AKPS score (7.5 points for QE and 7.2 points for HE), but the improvement did not reach the established minimal clinically important change threshold.

In Study 2, we found that PFP patients with self-reported signs of pain catastrophizing at baseline seemed to benefit more from QE than HE in the short term (12 weeks from baseline). Further, in the medium-term (26 weeks from baseline) participants with BMI above 25 kg/m^2 also seemed to benefit from QE, whereas patients with severe baseline knee pain seemed to benefit from HE. The results are subject to considerable uncertainty because of risk of multiplicity, small subgroups, and inconsistency over time.

In Study 3, we found that in patients with PFP, visual assessments of dynamic alignment during SLS and FL can be done reliably when the assessment is repeated by the same rater, i.e., moderate to good agreement. Further, we found moderate agreement when two experienced raters assessed the FL, while the interrater agreement was only 'fair' when assessing the SLS.

Our results from the RCT corroborate the results of a recently conducted systematic review with meta-analysis (17). Here it was concluded that isolated hip strengthening and knee strengthening were equivalent for the treatment of PFP, although it can be discussed, if the term equivalent is adequate, since all of the included studies were superiority studies. Further, the results are in line with the results from a recently published RCT (19). Here, 112 patients were randomized to a 6-week intervention consisting of patient education combined with either isolated hip-focused exercise, knee-focused exercise, or free physical activity. At three months follow-up there were no between-group differences for any of the primary (AKPS) or secondary (usual pain, worst pain, step-down, EQ-5D-5L, etc.) outcome measures at 3 months. The group as a whole improved with a mean of 7.6 points in AKPS, which is similar to the group improvement in our study, i.e., 7.5 points for QE and 7.2 points for HE. Another RCT supports the conclusion that hip focused exercises and quadriceps focused exercise provide similar results in terms of function, symptoms and pain (83), whereas others have found hip exercises to be superior to quadriceps exercises when comparing effectiveness (81, 82). Differences in population (females vs. mixed populations), insufficient sample sizes (ranging from 31 to 36

participants), and methodological limitations (no randomization and generally short intervention periods) make direct comparisons difficult.

Both QE and HE were associated with improvements in AKPS score, but the improvements did not reach the established minimally clinical important change threshold (8 points). The within-group changes for QE and HE are similar to those previously reported (19) but in the lower end compared to those reported in the other RCTs evaluating the effect of hip and knee exercises in adolescents and adults with PFP (83, 84, 101, 149-151). This difference may be explained by the settings of this study, where the mean pain duration was in the high end when compared to other interventional studies (83, 101, 151). Failure to report symptom duration in certain studies (84, 149) makes it difficult to compare study populations. Previous studies have shown that long symptom duration is associated with worse outcomes (irrespective of treatment) (6, 20) which may explain the somewhat smaller within group changes. Pain duration was not a significant effect modifier for HE or QE in our trial (Study 2), but the analyses cannot reveal if pain duration was a predictor for a successful outcome. The intention-to-treat analysis in Study 1 found equivalent between-group difference for AKPS, and thus, a secondary analysis where the whole group is treated as one cohort would be possible. This could give insights to whether pain duration (and other patient characteristics) are predictors of successful or unsuccessful outcomes.

Another plausible explanation is that the training interventions were home-based and with a minimum of interaction and supervision from health care providers, whereas comparable studies were primarily supervised. The impact of patients' expectations, emotions, and clinical context and the interaction between the patient and the therapist in clinical practice on health care outcomes is well established (152-154). All these factors can affect the treatment outcome and are linked to a placebo response (155). Supervised physical therapy has been found to produce greater improvements in symptom severity and physical function than unsupervised exercise in patients with lumbar spinal stenosis (156), potentially because of contextual factors. It is therefore plausible that the unsupervised setup of this study can explain the somewhat lower improvements seen. As with pain duration it is not possible to draw any conclusions on the effect of these psychosocial characteristics on treatment outcome based on present data.

The mechanisms by which strengthening exercises of the hip and quadriceps muscles elicit improvements in pain and function in patients with PFP are not fully understood. As stated in the introduction, one of the main theories is that patients with biomechanical deficits, e.g., functional knee valgus, improve as a consequence of "better" alignment. Dynamic knee valgus is associated with insufficient muscle strength in the hip abductors and external rotators (97, 98),

and therefore strengthening exercises for these muscles would theoretically improve malalignment. We did not find exercises for the hip muscles to be associated with a better outcome for patients with excessive knee valgus, and thus this biomechanical theory is not supported. Likewise, increments in hip muscle strength did not mediate the effect of hip exercises in a recently conducted trial (57), which further questions prevailing assumptions regarding the underlying mechanisms how hip exercises improve symptoms and function in patients with PFP.

Pain catastrophizing at baseline seemed to modify the 12-week treatment effect of QE vs HE, but the mechanisms can only be speculative. Pain catastrophizing is defined as a maladaptive cognitive-affective response to pain that involves negative thinking regarding the pain experience (157). It is a prevalent psychologic feature in patients with patellofemoral pain (158) and changes in maladaptive beliefs are associated with improved function and decreased pain (56). By graded exposure to exercises that put strain on the knees and incorporating patient education, one could hypothesize that pain beliefs could improve and potentially affect the symptoms and function (26). Pain catastrophizing was not an outcome measure and therefore this link is purely hypothetical. However, a recently proposed theoretical model for patients with anxiousness or fear of movement support graded exposure to knee loading activity (159). The influence of psychologic features including pain catastrophizing on treatment effects is highly prioritized in clinical practice (22) and research should elaborate on this in the future.

The mechanisms underlying 26-week treatment effects (BMI>25 benefit from QE and severe baseline knee pain benefit from HE) are also not possible to identify from the present data. Although high BMI has been linked to the development of PFP in adults and is a well-recognised risk factor for incidence and progression of knee OA (160, 161), no significant link has been established between BMI and intervention outcomes (162). An effect modifier analysis of a randomized controlled trial investigating the treatment effect of knee exercises compared to open-label placebo in individuals with knee OA did not find a high BMI to modify the effect (163). Intuitively, a high body mass would increase patellofemoral joint loads and stress during weight bearing exercises like squat and lunges (QE), contradictive of what present results suggest. The results underline that a biomechanical explanation for the effect of exercise is likely not exhaustive. Pain severity has been identified as an outcome predictor in PFP populations but not an effect modifier for specific treatments (20, 21). Severe cases often describe aggravated pain with activities that include weight bearing knee bends, such as running, stair climbing, squatting, and jumping (3). It therefore seems rational to avoid exercises like squats and lunges, at least in the beginning of a rehabilitation. As pain severity was only an effect modifier at week

26, gradual exposure to weight bearing quadriceps exercises after initial hip exercises may be recommendable, but again this is purely speculative.

In summary, we acknowledge that the identified potential treatment effect modifiers are merely indications that should be explored further. The absence of robust effect modifiers for PFP is in accordance with other studies comparing specific treatments. In a recently conducted RCT, patients with greater midfoot width mobility did not have superior benefits using foot orthoses compared to hip exercises at 12 weeks follow-up (164). Further, in a secondary analysis of an RCT comparing the effectiveness of supervised exercise therapy to usual care for 6 weeks in patients with PFP, none of the tested variables (sex, age, BMI, duration of complaints, and sports intensity) provided superior outcomes with treatment when comparing exercises for the quadriceps muscles combined with flexibility and balance exercises to a control intervention. Two factors, however, tended to have a predictive value in favor of exercise therapy: duration of complaints and sex (165).

The intrarater agreements for the SLS in the current study corroborate the results of a systematic review with meta-analysis elaborating on the inter- and intrarater agreement of visual assessment of the SLS including studies on both healthy subjects and subjects with lower extremity disorders (66). Here, the pooled results of the Kappa statistics showed a 'substantial' agreement for intrarater agreement (Kappa value 0.68 (95% CI 0.60 to 0.74)). Moreover, a 'moderate' agreement for interrater reliability of the SLS was found (Kappa value 0.58 (95% CI 0.50 to 0.65)), which is somewhat higher than in our study. In our study there were no training sessions in which the raters could synchronize assessments, to reflect a clinical setup where assessments are based on the experience of the assessors. This might explain the discrepancy when compared to results in the systematic review and the relatively low interrater agreements seen. For the FL, comparable intra- and inter agreement levels have been found in a recently conducted study (133). In agreement with our study, the study demonstrated higher levels of agreement for the FL compared to the SLS test.

Collectively, the clinical categorization of dynamic knee valgus is reliable (Study 3). As the mean difference between participants with and without dynamic knee malalignment was equal (Study 2), the relevance of such testing in the management of patients with PFP seems weakened.

Methodological considerations

In the primary analysis of the RCT, all participants were analyzed using the ITT population according to the intention-to-treat principle. Neither ITT nor Per Protocol (PP) analyses have

perfect properties in equivalence studies. Therefore, current recommendations state that both ITT and PP analysis should be done and support each other for equivalence to be claimed (114). The underlying principle is that when ITT and PP provide identical conclusions, the confidence for the study results is augmented. We choose the ITT as primary analysis because it preserves the advantages of randomization and is less prone to selection bias than PP. We tested the robustness of the results in the sensitivity analyses including PP analyses (table S1, Appendix 6).

The primary outcome measure (AKPS) was chosen because it is the most frequent used PROM in interventional studies specifically on PFP patients. The key secondary outcomes (change from baseline in the KOOS subscores 'pain', 'function' and 'quality of life' at week 12) were chosen because they are commonly used in clinical trials dealing with knee pain and because they supplement the items measured in AKPS. Other secondary outcomes were chosen to embrace the many impairments that people with PFP experience including quality of life, physical performance, and phycological features. Collectively, the results of the outcome measures included in this RCT can be compared to other interventional studies on PFP enabling an extension of current knowledge.

In Study 2, we included 20 patient characteristics that could potentially predict a superior outcome of one treatment over the other. Other studies limited the number of potential characteristics in order to minimize the risk of multiplicity (21, 165). Due to the explorative aim, we chose to include all the factors we could hypothesize to influence outcome. This have introduced a risk of false positive results and we are therefore humble in the interpretation and conclusion of our results.

In study 3 we chose a pragmatic way of assessing the agreement of visually assessing the alignment of patients with PFP. By recording and saving the performed movements on an Ipad, we lose the means of assessing the intervariability of the subjects. On the other hand, this allows us to focus on the raters' ability to reproduce the judgements without the 'noise' of performance variability.

Limitations

There are some limitations to the studies embraced in this PhD, that need acknowledgement. First, the exercise programs were home-based with limited supervision which may introduce a risk that the exercises were not performed correctly. This can, however, also be seen as a strength, as it resembles a typical clinical setup where high degree of supervision normally is unfeasible (22). Most of the interventions in comparable studies were supervised, but this is not always feasible in a clinical setting. The patient–physiotherapist relationship and the overall

healthcare setting are relevant categories of contextual factors that may lead to a biased overall treatment effect (152). Further, the exercise adherence data was based on self-reporting, which introduces an inherent risks of overestimation due to social desirability, recall period, and selective recall (166). A limitation that also needs consideration is the single-center setup which potentially limits the external validity. By incorporating other hospitals in the study organization, we could have secured a wider representation of patients, e.g., from other parts of the country (or world). Finally, as part of Study 2 the physiotherapists recorded the participants' projected prognosis after the first clinical encounter (the reporting of this study will be done in a separate manuscript) possibly introducing expectation bias on the part of the physiotherapists. However, the act of explicitly stating an estimation of prognosis is not different from the implicit, unconscious estimation that physiotherapists (and other clinicians) do automatically when seeing a patient.

Limitations to Study 3 include the 2-dimensional setup that potentially excluded vital components of movement quality, e.g., axial plane control of the lower extremity. Further, we focused on the knee excursion *per se* and did thereby exclude other body segments like the pelvis or the foot.

Strengths

One of the main strengths of the main RCT is the equivalence design allowing us to draw adequate conclusions on the comparative effectiveness of hip focused and quadriceps focused exercises. The equivalence analysis is based on a clinically meaningful difference and not merely on statistical significance (p-values) (123). This makes the interpretation of results clinically meaningful and heightens the precision of the estimated differences. Another prominent strength that should be highlighted is the large number of participants included in Study 1 and 2. The sample size was conservatively set based on a standard deviation of 15 in AKPS, which is in the high end of the wide range reported in comparable trials (149, 167). The anticipated dropout rate was also set conservatively to 20%. The strength of high sample sizes also relates to Study 3, which strengthens the external validity. Further, the statistical analysis plans that were signed and submitted to clinicaltrials.gov before manuscript submission is also a strength (Appendix 5 and 6). Finally, the intervention period of 12 weeks in the main trial is higher than comparable interventional studies (17, 168).

Clinical implications

Rehabilitation consisting of quadriceps focused exercises and hip muscles focused exercises elicit equal improvements in symptoms and function. As a consequence, clinicians should use the patients' preferences and own clinical reasoning when deciding on treatment. Shared decision making may improve healthcare efficiency and is recommended in the rehabilitation of patients with PFP (11, 169, 170). Further, no robust effect modifier was found, leaving the clinician with few guidelines when deciding on the optimal treatment for their patients. Clinicians may focus more or less on exercises for the quadriceps or hips according to the potential effect modifiers found in the present study (acknowledging the uncertainty of the findings) and patient preferences. The identification of the subgroups (signs of pain catastrophizing, overweight, and pain severity) is clinically feasible as they rely on simple patient reported outcomes and measurements of height and weight.

This thesis does not support a purely biomechanical approach to the assessment or the treatment of patients with patellofemoral pain. Focusing on biomechanical deficits may even reinforce negative psychosocial features of PFP like pain-related fear, catastrophising and pain self-efficacy, which are associated with pain and disability (55, 159). A pathomechanical framework may still be part of the puzzle, but clinicians also need to consider psychological and social factors (159, 171).

Conclusions

In a population of PFP patients, we found that a quadriceps focused and a hip muscles focused exercise program yielded equivalent effectiveness on symptoms and function in both short and medium term (12 and 26 weeks). Moreover, we found indications that patients with signs of pain catastrophizing and high BMI benefit more from the quadriceps focused exercises (vs hip focused exercise), whereas patients with severe pain benefit from hip muscles focused exercises (vs quadriceps focused exercise). The differential treatment responses are imprecisely estimated, and the identified potential treatment effect modifiers are merely indications that should be explored further. Finally, we found that visual assessment of a forward lunge (as part of the effect modifier study) can be done reliably, whereas the assessment of a single leg squat should preferably be performed and repeated by the same tester.

Perspectives

For decades, researchers have investigated the effect of different treatment modalities in patients suffering from PFP. It seems the optimal treatment has not been found. Patients suffer for years and decades despite undergoing well-intentioned and -conducted rehabilitation (5, 6, 109). One major problem with trying to find the optimal treatment is that the pathogenesis and pathophysiology are unknown. More specifically, we do not know whether the pain comes from specific tissues or if the pain is non-specific and driven by sensitization or other pain mechanisms. There are current endeavors aiming to assess various specific mechanisms including subchondral bone metabolism (ClinicalTrials.gov Identifier: NCT03784235). Results from such studies may open new avenues for diagnosis and treatment in the future.

Today, the recommended management approach is to focus on knee and hip exercise, and education to balance rest and activity. This treatment often requires many visits with a therapist, and it is hypothesized that part of the positive effect of this approach might be contextual and not due to tissue specific effects. With the present contribution to the field of PFP research, we question the pursuing of biomechanics or specific tissue as drivers of the disorder and as the primary aim of rehabilitation. In a recently held patellofemoral pain summit meeting including international clinicians and researchers who are active in the field of patellofemoral pain, it was suggested that interventions could be based more on patient education, encouragement of self-management, and reducing the long-term negative consequences of patellofemoral pain: increases in BMI, sedentary behavior and increases in anxiety and depression (unpublished). We are in the middle of new departures in chronic pain research, and traditional theories that have prevailed for decades are being tossed in the air. As a scientist it is important to accept the consequences of the research and to adjust the theory according to the data, rather than trying to make the data fit the theory. Hopefully, by doing that, research will lead the way to more effective treatment for this patient group in the future.

References

1. Mousavi SH, Hijmans JM, Minoonejad H, Rajabi R, Zwerver J. Factors Associated With Lower Limb Injuries in Recreational Runners: A Cross-Sectional Survey Including Mental Aspects and Sleep Quality. J Sports Sci Med. 2021;20(2):204-15.
2. Smith BE, Selfe J, Thacker D, Hendrick P, Bateman M, Moffatt F, et al. Incidence and prevalence of patellofemoral pain: A systematic review and meta-analysis. PLoS One. 2018;13(1):e0190892.
3. Crossley KM, Stefanik JJ, Selfe J, Collins NJ, Davis IS, Powers CM, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br J Sports Med. 2016;50(14):839-43.
4. Rathleff CR, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Half of 12-15-year-olds with knee pain still have pain after one year. Dan Med J. 2013;60(11):A4725.
5. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain. Am J Sports Med. 2016;44(5):1165-71.
6. Lankhorst NE, van Middelkoop M, Crossley KM, Bierma-Zeinstra SM, Oei EH, Vicenzino B, et al. Factors that predict a poor outcome 5-8 years after the diagnosis of patellofemoral pain: a multicentre observational analysis. Br J Sports Med. 2016;50(14):881-6.
7. Andreucci A, Roos EM, Rasmussen S, Olesen JL, Holmich P, Thorborg K, et al. Analgesic use in adolescents with patellofemoral pain or Osgood-Schlatter Disease: a secondary cross-sectional analysis of 323 subjects. Scand J Pain. 2021.
8. Lack S, Neal B, De Oliveira Silva D, Barton C. How to manage patellofemoral pain - Understanding the multifactorial nature and treatment options. Phys Ther Sport. 2018;32:155-66.
9. Crossley KM, van Middelkoop M, Barton CJ, Culvenor AG. Rethinking patellofemoral pain: Prevention, management and long-term consequences. Best Pract Res Clin Rheumatol. 2019;33(1):48-65.
10. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'Best Practice Guide to Conservative Management of Patellofemoral Pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med. 2015;49(14):923-34.
11. Collins NJ, Barton CJ, van Middelkoop M, Callaghan MJ, Rathleff MS, Vicenzino BT, et al. 2018 Consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th International Patellofemoral Pain Research Retreat, Gold Coast, Australia, 2017. Br J Sports Med. 2018;52(18):1170-8.
12. Kooiker L, Van De Port IG, Weir A, Moen MH. Effects of physical therapist-guided quadriceps-strengthening exercises for the treatment of patellofemoral pain syndrome: a systematic review. J Orthop Sports Phys Ther. 2014;44(6):391-B1.
13. Petersen W, Ellermann A, Gosele-Koppenburg A, Best R, Rembitzki IV, Bruggemann GP, et al. Patellofemoral pain syndrome. Knee Surg Sports Traumatol Arthrosc. 2014;22(10):2264-74.
14. Santos TR, Oliveira BA, Ocarino JM, Holt KG, Fonseca ST. Effectiveness of hip muscle strengthening in patellofemoral pain syndrome patients: a systematic review. Braz J Phys Ther. 2015;19(3):167-76.

15. Burrows NJ, Booth J, Sturnieks DL, Barry BK. Acute resistance exercise and pressure pain sensitivity in knee osteoarthritis: a randomised crossover trial. Osteoarthritis Cartilage. 2014;22(3):407-14.
16. Vaegter HB, Jones MD. Exercise-induced hypoalgesia after acute and regular exercise: experimental and clinical manifestations and possible mechanisms in individuals with and without pain. Pain Rep. 2020;5(5):e823.
17. Na Y, Han C, Shi Y, Zhu Y, Ren Y, Liu W. Is Isolated Hip Strengthening or Traditional Knee-Based Strengthening More Effective in Patients With Patellofemoral Pain Syndrome? A Systematic Review With Meta-analysis. Orthop J Sports Med. 2021;9(7):23259671211017503.
18. Saltychev M, Dutton RA, Laimi K, Beaupre GS, Virolainen P, Fredericson M. Effectiveness of conservative treatment for patellofemoral pain syndrome: A systematic review and meta-analysis. J Rehabil Med. 2018;50(5):393-401.
19. Hott A, Brox JI, Pripp AH, Juel NG, Paulsen G, Liavaag S. Effectiveness of Isolated Hip Exercise, Knee Exercise, or Free Physical Activity for Patellofemoral Pain: A Randomized Controlled Trial. Am J Sports Med. 2019;47(6):1312-22.
20. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley K, et al. Can we predict the outcome for people with patellofemoral pain? A systematic review on prognostic factors and treatment effect modifiers. Br J Sports Med. 2017;51(23):1650-60.
21. Hott A, Brox JI, Pripp AH, Juel NG, Liavaag S. Predictors of Pain, Function, and Change in Patellofemoral Pain. Am J Sports Med. 2020;48(2):351-8.
22. Smith BE, Hendrick P, Bateman M, Moffatt F, Rathleff MS, Selfe J, et al. Current management strategies for patellofemoral pain: an online survey of 99 practising UK physiotherapists. BMC Musculoskelet Disord. 2017;18(1):181.
23. Jensen HAR, Davidsen M, Møller SR, Román JEI, Kragelund K, Christensen AI, et al. Danskernes sundhed – Den Nationale Sundhedsprofil 2021. 2022 10.03.2022.
24. Larsen JB. ENHEDEN FOR MUSKULOSKELETAL SUNDHED [Available from: https://www.almenmedicin.aau.dk/forskning/center-for-musculoskeletal-health/painpath/enheden-muskuloskeletal-sundhed/.
25. Hryvniak D, Magrum E, R. W. Patellofemoral Pain Syndrome: An Update. Curr Phys Med Rehabil Rep. 2014;2:16-24.
26. Wood L, Muller S, Peat G. The epidemiology of patellofemoral disorders in adulthood: a review of routine general practice morbidity recording. Prim Health Care Res Dev. 2011;12(2):157-64.
27. Roush JR, Curtis Bay R. Prevalence of anterior knee pain in 18-35 year-old females. Int J Sports Phys Ther. 2012;7(4):396-401.
28. Martinez-Cano JP, Ramos-Rivera JC, Gomez-Garcia J, Casas-Barragan GA, Rosales MC, Escobar-Gonzalez SS. Anterior knee pain in runners after a half-marathon race. J Clin Orthop Trauma. 2021;23:101640.
29. Satterthwaite P, Norton R, Larmer P, Robinson E. Risk factors for injuries and other health problems sustained in a marathon. Br J Sports Med. 1999;33(1):22-6.
30. Rathleff MS, Vicenzino B, Middelkoop M, Graven-Nielsen T, van Linschoten R, Holmich P, et al. Patellofemoral Pain in Adolescence and Adulthood: Same Same, but Different? Sports Med. 2015;45(11):1489-95.
31. D'Ambrosi R, Meena A, Raj A, Ursino N, Hewett TE. Anterior Knee Pain: State of the Art. Sports Med Open. 2022;8(1):98.
32. Crossley KM. Is patellofemoral osteoarthritis a common sequela of patellofemoral pain? Br J Sports Med. 2014;48(6):409-10.

33. Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. J Orthop Sports Phys Ther. 2003;33(11):639-46.
34. Dye SF. The pathophysiology of patellofemoral pain: a tissue homeostasis perspective. Clin Orthop Relat Res. 2005(436):100-10.
35. MacIntyre NJ, Hill NA, Fellows RA, Ellis RE, Wilson DR. Patellofemoral joint kinematics in individuals with and without patellofemoral pain syndrome. J Bone Joint Surg Am. 2006;88(12):2596-605.
36. Fick CN, Grant C, Sheehan FT. Patellofemoral Pain in Adolescents: Understanding Patellofemoral Morphology and Its Relationship to Maltracking. Am J Sports Med. 2020;48(2):341-50.
37. Shen A, Boden BP, Grant C, Carlson VR, Alter KE, Sheehan FT. Adolescents and adults with patellofemoral pain exhibit distinct patellar maltracking patterns. Clin Biomech (Bristol, Avon). 2021;90:105481.
38. Ferrari D, Briani RV, de Oliveira Silva D, Pazzinatto MF, Ferreira AS, Alves N, et al. Higher pain level and lower functional capacity are associated with the number of altered kinematics in women with patellofemoral pain. Gait Posture. 2018;60:268-72.
39. Ashnagar Z, Hadian MR, Sajjadi E, Kajbafvala M, Olyaei G, Pashazadeh F, et al. Quadriceps architecture in individuals with patellofemoral pain: A systematic review. J Bodyw Mov Ther. 2021;25:248-54.
40. Ducatti MHM, Waiteman MC, Botta AFB, Lopes HDS, Glaviano NR, Azevedo FM, et al. Knee flexor strength, rate of torque development and flexibility in women and men with patellofemoral pain: Relationship with pain and the performance in the single leg bridge test. Phys Ther Sport. 2021;50:166-72.
41. Neal BS, Barton CJ, Gallie R, O'Halloran P, Morrissey D. Runners with patellofemoral pain have altered biomechanics which targeted interventions can modify: A systematic review and meta-analysis. Gait Posture. 2016;45:69-82.
42. Lankhorst NE, Bierma-Zeinstra SM, van Middelkoop M. Factors associated with patellofemoral pain syndrome: a systematic review. Br J Sports Med. 2013;47(4):193-206.
43. Thomas MJ, Wood L, Selfe J, Peat G. Anterior knee pain in younger adults as a precursor to subsequent patellofemoral osteoarthritis: a systematic review. BMC Musculoskelet Disord. 2010;11:201.
44. Wyndow N, Collins N, Vicenzino B, Tucker K, Crossley K. Is There a Biomechanical Link Between Patellofemoral Pain and Osteoarthritis? A Narrative Review. Sports Med. 2016;46(12):1797-808.
45. Utting MR, Davies G, Newman JH. Is anterior knee pain a predisposing factor to patellofemoral osteoarthritis? Knee. 2005;12(5):362-5.
46. Hinman RS, Lentzos J, Vicenzino B, Crossley KM. Is patellofemoral osteoarthritis common in middle-aged people with chronic patellofemoral pain? Arthritis Care Res (Hoboken). 2014;66(8):1252-7.
47. Draper CE, Fredericson M, Gold GE, Besier TF, Delp SL, Beaupre GS, et al. Patients with patellofemoral pain exhibit elevated bone metabolic activity at the patellofemoral joint. J Orthop Res. 2012;30(2):209-13.
48. https://www.iasp-pain.org/terminology IAftSoPWs. IASP terminology. Accessed September 5, 2022. [
49. Pazzinatto MF, de Oliveira Silva D, Barton C, Rathleff MS, Briani RV, de Azevedo FM. Female Adults with Patellofemoral Pain Are Characterized by Widespread Hyperalgesia, Which Is Not Affected Immediately by Patellofemoral Joint Loading. Pain Med. 2016;17(10):1953-61.

50. Rathleff MS, Petersen KK, Arendt-Nielsen L, Thorborg K, Graven-Nielsen T. Impaired Conditioned Pain Modulation in Young Female Adults with Long-Standing Patellofemoral Pain: A Single Blinded Cross-Sectional Study. Pain Med. 2016;17(5):980-8.
51. Boudreau SA, Kamavuako EN, Rathleff MS. Distribution and symmetrical patellofemoral pain patterns as revealed by high-resolution 3D body mapping: a cross-sectional study. BMC Musculoskelet Disord. 2017;18(1):160.
52. Holden S, Straszek CL, Rathleff MS, Petersen KK, Roos EM, Graven-Nielsen T. Young females with long-standing patellofemoral pain display impaired conditioned pain modulation, increased temporal summation of pain, and widespread hyperalgesia. Pain. 2018;159(12):2530-7.
53. Arendt-Nielsen L, Nie H, Laursen MB, Laursen BS, Madeleine P, Simonsen OH, et al. Sensitization in patients with painful knee osteoarthritis. Pain. 2010;149(3):573-81.
54. Arendt-Nielsen L, Graven-Nielsen T. Translational musculoskeletal pain research. Best Pract Res Clin Rheumatol. 2011;25(2):209-26.
55. Maclachlan LR, Collins NJ, Matthews MLG, Hodges PW, Vicenzino B. The psychological features of patellofemoral pain: a systematic review. Br J Sports Med. 2017;51(9):732-42.
56. Selhorst M, Fernandez-Fernandez A, Schmitt L, Hoehn J. Effect of a Psychologically Informed Intervention to Treat Adolescents With Patellofemoral Pain: A Randomized Controlled Trial. Arch Phys Med Rehabil. 2021;102(7):1267-73.
57. Holden S, Matthews M, Rathleff MS, Kasza J, Fohx G, Vicenzino B. How Do Hip Exercises Improve Pain in Individuals With Patellofemoral Pain? Secondary Mediation Analysis of Strength and Psychological Factors as Mechanisms. J Orthop Sports Phys Ther. 2021;51(12):602-10.
58. Vicenzino BT, Rathleff MS, Holden S, Maclachlan L, Smith BE, de Oliveira Silva D, et al. Developing Clinical and Research Priorities for Pain and Psychological Features in People With Patellofemoral Pain: An International Consensus Process With Health Care Professionals. J Orthop Sports Phys Ther. 2022;52(1):29-39.
59. Nunes GS, Stapait EL, Kirsten MH, de Noronha M, Santos GM. Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis. Phys Ther Sport. 2013;14(1):54-9.
60. Cook C, Mabry L, Reiman MP, Hegedus EJ. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. Physiotherapy. 2012;98(2):93-100.
61. Cook C, Hegedus E, Hawkins R, Scovell F, Wyland D. Diagnostic accuracy and association to disability of clinical test findings associated with patellofemoral pain syndrome. Physiother Can. 2010;62(1):17-24.
62. Messier SP, Davis SE, Curl WW, Lowery RB, Pack RJ. Etiologic factors associated with patellofemoral pain in runners. Med Sci Sports Exerc. 1991;23(9):1008-15.
63. Powers CM, Witvrouw E, Davis IS, Crossley KM. Evidence-based framework for a pathomechanical model of patellofemoral pain: 2017 patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester, UK: part 3. Br J Sports Med. 2017;51(24):1713-23.
64. Neal BS, Lack SD, Lankhorst NE, Raye A, Morrissey D, van Middelkoop M. Risk factors for patellofemoral pain: a systematic review and meta-analysis. Br J Sports Med. 2019;53(5):270-81.
65. Hansen R, Lundgaard-Nielsen M, Henriksen M. Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study. PeerJ. 2021;9:e12203.
66. Ressman J, Grooten WJA, Rasmussen Barr E. Visual assessment of movement quality in the single leg squat test: a review and meta-analysis of inter-rater and intrarater reliability. BMJ Open Sport Exerc Med. 2019;5(1):e000541.

67. Gwynne CR, Curran SA. Two-dimensional frontal plane projection angle can identify subgroups of patellofemoral pain patients who demonstrate dynamic knee valgus. Clin Biomech (Bristol, Avon). 2018;58:44-8.
68. Willson JD, Davis IS. Utility of the frontal plane projection angle in females with patellofemoral pain. J Orthop Sports Phys Ther. 2008;38(10):606-15.
69. Whatman C, Hing W, Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. Phys Ther Sport. 2012;13(2):87-96.
70. Whatman C, Hume P, Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Phys Ther Sport. 2013;14(3):168-74.
71. Piva SR, Fitzgerald GK, Wisniewski S, Delitto A. Predictors of pain and function outcome after rehabilitation in patients with patellofemoral pain syndrome. J Rehabil Med. 2009;41(8):604-12.
72. Arendt-Nielsen L, Morlion B, Perrot S, Dahan A, Dickenson A, Kress HG, et al. Assessment and manifestation of central sensitisation across different chronic pain conditions. Eur J Pain. 2018;22(2):216-41.
73. Hasan S, Alonazi A, Anwer S, Jamal A, Parvez S, Alfaiz FAS, et al. Efficacy of Patellar Taping and Electromyographic Biofeedback Training at Various Knee Angles on Quadriceps Strength and Functional Performance in Young Adult Male Athletes with Patellofemoral Pain Syndrome: A Randomized Controlled Trial. Pain Res Manag. 2022;2022:8717932.
74. Karamiani F, Mostamand J, Rahimi A, Nasirian M. The Effect of Gluteus Medius Dry Needling on Pain and Physical Function of Non-athlete women with Unilateral Patellofemoral Pain Syndrome: A Double-Blind Randomized Clinical Trial. J Bodyw Mov Ther. 2022;30:23-9.
75. Pocai BL, Provensi E, Serighelli F, Rigo G, Artioli DP, de Albuquerque CE, et al. Effect of photobiomodulation in the patellofemoral pain syndrome; randomized clinical trial in young women. J Bodyw Mov Ther. 2021;26:263-7.
76. Shadloo N, Kamali F, Salehi Dehno N. A comparison between whole-body vibration and conventional training on pain and performance in athletes with patellofemoral pain. J Bodyw Mov Ther. 2021;27:661-6.
77. Fanlo-Mazas P, Bueno-Gracia E, Ruiz de Escudero-Zapico A, Lopez-de-Celis C, Hidalgo-Garcia C, Rodriguez-Sanz J, et al. The Effect of Diacutaneous Fibrolysis on Local and Widespread Hyperalgesia and Muscle Length in Patients With Patellofemoral Pain Syndrome: Secondary Analysis of a Pretest-Posttest Clinical Trial. J Sport Rehabil. 2021;30(5):804-11.
78. Winters M, Holden S, Lura CB, Welton NJ, Caldwell DM, Vicenzino BT, et al. Comparative effectiveness of treatments for patellofemoral pain: a living systematic review with network meta-analysis. Br J Sports Med. 2020.
79. Lack S, Barton C, Sohan O, Crossley K, Morrissey D. Proximal muscle rehabilitation is effective for patellofemoral pain: a systematic review with meta-analysis. Br J Sports Med. 2015;49(21):1365-76.
80. Neal BS, Bartholomew C, Barton CJ, Morrissey D, Lack SD. Six treatments have positive effects at three-months for people with patellofemoral pain: a systematic review with meta-analysis. J Orthop Sports Phys Ther. 2022:1-52.
81. Dolak KL, Silkman C, Medina McKeon J, Hosey RG, Lattermann C, Uhl TL. Hip strengthening prior to functional exercises reduces pain sooner than quadriceps strengthening in females with patellofemoral pain syndrome: a randomized clinical trial. J Orthop Sports Phys Ther. 2011;41(8):560-70.

82. Khayambashi K, Fallah A, Movahedi A, Bagwell J, Powers C. Posterolateral hip muscle strengthening versus quadriceps strengthening for patellofemoral pain: a comparative control trial. Arch Phys Med Rehabil. 2014;95(5):900-7.
83. Ferber R, Bolgla L, Earl-Boehm JE, Emery C, Hamstra-Wright K. Strengthening of the hip and core versus knee muscles for the treatment of patellofemoral pain: a multicenter randomized controlled trial. J Athl Train. 2015;50(4):366-77.
84. Saad MC, Vasconcelos RA, Mancinelli LVO, Munno MSB, Liporaci RF, Grossi DB. Is hip strengthening the best treatment option for females with patellofemoral pain? A randomized controlled trial of three different types of exercises. Braz J Phys Ther. 2018;22(5):408-16.
85. Kami K, Tajima F, Senba E. Brain Mechanisms of Exercise-Induced Hypoalgesia: To Find a Way Out from "Fear-Avoidance Belief". Int J Mol Sci. 2022;23(5).
86. Kraemer WJ, Ratamess NA, French DN. Resistance training for health and performance. Curr Sports Med Rep. 2002;1(3):165-71.
87. Bolgla LA, Earl-Boehm J, Emery C, Hamstra-Wright K, Ferber R. Comparison of hip and knee strength in males with and without patellofemoral pain. Phys Ther Sport. 2015;16(3):215-21.
88. Lankhorst NE, Bierma-Zeinstra SM, van Middelkoop M. Risk factors for patellofemoral pain syndrome: a systematic review. J Orthop Sports Phys Ther. 2012;42(2):81-94.
89. Hart JM, Pietrosimone B, Hertel J, Ingersoll CD. Quadriceps activation following knee injuries: a systematic review. J Athl Train. 2010;45(1):87-97.
90. Rio E, van Ark M, Docking S, Moseley GL, Kidgell D, Gaida JE, et al. Isometric Contractions Are More Analgesic Than Isotonic Contractions for Patellar Tendon Pain: An In-Season Randomized Clinical Trial. Clin J Sport Med. 2017;27(3):253-9.
91. Vaegter HB. Exercising non-painful muscles can induce hypoalgesia in individuals with chronic pain. Scand J Pain. 2017;15:60-1.
92. Rice D, Nijs J, Kosek E, Wideman T, Hasenbring MI, Koltyn K, et al. Exercise-Induced Hypoalgesia in Pain-Free and Chronic Pain Populations: State of the Art and Future Directions. J Pain. 2019;20(11):1249-66.
93. Bolgla LA, Malone TR, Umberger BR, Uhl TL. Comparison of hip and knee strength and neuromuscular activity in subjects with and without patellofemoral pain syndrome. Int J Sports Phys Ther. 2011;6(4):285-96.
94. Prins MR, van der Wurff P. Females with patellofemoral pain syndrome have weak hip muscles: a systematic review. Aust J Physiother. 2009;55(1):9-15.
95. McKenzie K, Galea V, Wessel J, Pierrynowski M. Lower extremity kinematics of females with patellofemoral pain syndrome while stair stepping. J Orthop Sports Phys Ther. 2010;40(10):625-32.
96. Wirtz AD, Willson JD, Kernozek TW, Hong DA. Patellofemoral joint stress during running in females with and without patellofemoral pain. Knee. 2012;19(5):703-8.
97. Powers CM. The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. J Orthop Sports Phys Ther. 2010;40(2):42-51.
98. Rathleff MS, Rathleff CR, Crossley KM, Barton CJ. Is hip strength a risk factor for patellofemoral pain? A systematic review and meta-analysis. Br J Sports Med. 2014;48(14):1088.
99. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med. 2016.

100. Baldon Rde M, Serrao FV, Scattone Silva R, Piva SR. Effects of functional stabilization training on pain, function, and lower extremity biomechanics in women with patellofemoral pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2014;44(4):240-A8.
101. Fukuda TY, Melo WP, Zaffalon BM, Rossetto FM, Magalhaes E, Bryk FF, et al. Hip posterolateral musculature strengthening in sedentary women with patellofemoral pain syndrome: a randomized controlled clinical trial with 1-year follow-up. J Orthop Sports Phys Ther. 2012;42(10):823-30.
102. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. J Clin Epidemiol. 2021;134:174-7.
103. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther. 2009;89(7):698-704.
104. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. JAMA. 2006;296(10):1286-9.
105. Selfe J, Janssen J, Callaghan M, Witvrouw E, Sutton C, Richards J, et al. Are there three main subgroups within the patellofemoral pain population? A detailed characterisation study of 127 patients to help develop targeted intervention (TIPPs). Br J Sports Med. 2016;50(14):873-80.
106. Selhorst M, Rice W, Degenhart T, Jackowski M, Tatman M. Evaluation of a treatment algorithm for patients with patellofemoral pain syndrome: a pilot study. Int J Sports Phys Ther. 2015;10(2):178-88.
107. Lack S, Barton C, Vicenzino B, Morrissey D. Outcome predictors for conservative patellofemoral pain management: a systematic review and meta-analysis. Sports Med. 2014;44(12):1703-16.
108. Blond L, Hansen L. Patellofemoral pain syndrome in athletes: a 5.7-year retrospective follow-up study of 250 athletes. Acta Orthop Belg. 1998;64(4):393-400.
109. Collins NJ, Bierma-Zeinstra SM, Crossley KM, van Linschoten RL, Vicenzino B, van Middelkoop M. Prognostic factors for patellofemoral pain: a multicentre observational analysis. Br J Sports Med. 2013;47(4):227-33.
110. Collins NJ, Crossley KM, Darnell R, Vicenzino B. Predictors of short and long term outcome in patellofemoral pain syndrome: a prospective longitudinal study. BMC Musculoskelet Disord. 2010;11:11.
111. Natri A, Kannus P, Jarvinen M. Which factors predict the long-term outcome in chronic patellofemoral pain syndrome? A 7-yr prospective follow-up study. Med Sci Sports Exerc. 1998;30(11):1572-7.
112. Kraemer HC. Messages for Clinicians: Moderators and Mediators of Treatment Outcome in Randomized Clinical Trials. Am J Psychiatry. 2016;173(7):672-9.
113. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? Arch Phys Med Rehabil. 2004;85(5):815-22.
114. Johnston BC PD, Devji T, Maxwell LJ, Bingham III CO, Beaton D, Boers M, Briel M, Busse JW, Carrasco-Labra A, Christensen R, da Costa BR, El Dib R, Lyddiatt A, Ostelo RW, Shea B, Singh J, Terwee CB, Williamson PR, Gagnier JJ, Tugwell P, Guyatt GH. Chapter 18: Patient-reported outcomes. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook. [
115. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. BMJ Evid Based Med. 2022;27(3):153-5.
116. Esculier JF, Roy JS, Bouyer LJ. Psychometric evidence of self-reported questionnaires for patellofemoral pain syndrome: a systematic review. Disabil Rehabil. 2013;35(26):2181-90.

117. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy. 1993;9(2):159-63.
118. Crossley KM, Macri EM, Cowan SM, Collins NJ, Roos EM. The patellofemoral pain and osteoarthritis subscale of the KOOS (KOOS-PF): development and validation using the COSMIN checklist. Br J Sports Med. 2018;52(17):1130-6.
119. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther. 1998;28(2):88-96.
120. Garrison C, Cook C. Clinimetrics corner: the Global Rating of Change Score (GRoC) poorly correlates with functional measures and is not temporally stable. J Man Manip Ther. 2012;20(4):178-81.
121. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018;125(13):1716.
122. Stefanos R, Graziella D, Giovanni T. Methodological aspects of superiority, equivalence, and non-inferiority trials. Intern Emerg Med. 2020;15(6):1085-91.
123. Kishore K, Mahajan R. Understanding Superiority, Noninferiority, and Equivalence for Clinical Trials. Indian Dermatol Online J. 2020;11(6):890-4.
124. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. Theriogenology. 2010;73(9):1167-79.
125. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. J Hum Reprod Sci. 2011;4(1):8-11.
126. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. Trials. 2015;16:405.
127. Broglio K. Randomization in Clinical Trials: Permuted Blocks and Stratification. JAMA. 2018;319(21):2223-4.
128. Randelli P, Arrigoni P, Lubowitz JH, Cabitza P, Denti M. Randomization procedures in orthopaedic trials. Arthroscopy. 2008;24(7):834-8.
129. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud. 2011;48(6):661-71.
130. Holden S, Rathleff MS, Jensen MB, Barton CJ. How can we implement exercise therapy for patellofemoral pain if we don't know what was prescribed? A systematic review. Br J Sports Med. 2018;52(6):385.
131. Toigo M, Boutellier U. New fundamental resistance exercise determinants of molecular and cellular muscle adaptations. Eur J Appl Physiol. 2006;97(6):643-63.
132. American College of Sports M. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.
133. Whelan D, Delahunt E, O'Reilly M, Hernandez B, Caulfield B. Determining Interrater and Intrarater Levels of Agreement in Students and Clinicians When Visually Evaluating Movement Proficiency During Screening Assessments. Phys Ther. 2019;99(4):478-86.
134. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. Scand J Med Sci Sports. 2010;20(3):493-501.
135. Romero-Franco N, Jimenez-Reyes P, Montano-Munuera JA. Validity and reliability of a low-cost digital dynamometer for measuring isometric strength of lower limb. J Sports Sci. 2017;35(22):2179-84.

136. Mentiplay BF, Perraton LG, Bower KJ, Adair B, Pua YH, Williams GP, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. PLoS One. 2015;10(10):e0140822.
137. Klokker L, Christensen R, Waehrens EE, Bandak E, Bartholdy C, Bliddal H, et al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: construct validity, responsiveness, and interpretability in a research setting. Health Qual Life Outcomes. 2016;14:91.
138. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain. 2007;11(2):153-63.
139. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33(5):337-43.
140. Grahame R, Bird HA, Child A. The revised (Brighton 1998) criteria for the diagnosis of benign joint hypermobility syndrome (BJHS). J Rheumatol. 2000;27(7):1777-9.
141. Folci M, Capsoni F. Arthralgias, fatigue, paresthesias and visceral pain: can joint hypermobility solve the puzzle? A case report. BMC Musculoskelet Disord. 2016;17:58.
142. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22(10):1911-20.
143. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med. 2000;23(4):351-65.
144. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. J Behav Med. 1997;20(6):589-605.
145. Adams JH, Williams AC. What affects return to work for graduates of a pain management program with chronic upper limb pain? J Occup Rehabil. 2003;13(2):91-106.
146. McPoil TG, Vicenzino B, Cornwall MW, Collins N, Warren M. Reliability and normative values for the foot mobility magnitude: a composite measure of vertical and medial-lateral mobility of the midfoot. J Foot Ankle Res. 2009;2:6.
147. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
148. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
149. Fukuda TY, Rossetto FM, Magalhaes E, Bryk FF, Lucareli PR, de Almeida Aparecida Carvalho N. Short-term effects of hip abductors and lateral rotators strengthening in females with patellofemoral pain syndrome: a randomized controlled clinical trial. J Orthop Sports Phys Ther. 2010;40(11):736-42.
150. Almeida GPL, Rodrigues H, Coelho BAL, Rodrigues CAS, Lima POP. Anteromedial versus posterolateral hip musculature strengthening with dose-controlled in women with patellofemoral pain: A randomized controlled trial. Phys Ther Sport. 2021;49:149-56.
151. Constantinou A, Mamais I, Papathanasiou G, Lamnisos D, Stasinopoulos D. Comparing hip and knee focused exercises versus hip and knee focused exercises with the use of blood flow restriction training in adults with patellofemoral pain. Eur J Phys Rehabil Med. 2022;58(2):225-35.
152. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet. 2001;357(9258):757-62.
153. Hafliethadottir SH, Juhl CB, Nielsen SM, Henriksen M, Harris IA, Bliddal H, et al. Placebo response and effect in randomized clinical trials: meta-research with focus on contextual effects. Trials. 2021;22(1):493.

154. Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. BMC Musculoskelet Disord. 2018;19(1):27.
155. Rossettini G, Palese A, Geri T, Mirandola M, Tortella F, Testa M. The Knowledge of Contextual Factors as Triggers of Placebo and Nocebo Effects in Patients With Musculoskeletal Pain: Findings From a National Survey. Front Psychiatry. 2019;10:478.
156. Minetama M, Kawakami M, Teraguchi M, Kagotani R, Mera Y, Sumiya T, et al. Supervised physical therapy versus unsupervised exercise for patients with lumbar spinal stenosis: 1-year follow-up of a randomized controlled trial. Clin Rehabil. 2021;35(7):964-75.
157. Sullivan MJL, Lynch ME, Clark AJ. Dimensions of catastrophic thinking associated with pain experience and disability in patients with neuropathic pain conditions. Pain. 2005;113(3):310-5.
158. Glaviano NR, Holden S, Bazett-Jones DM, Singe SM, Rathleff MS. Living well (or not) with patellofemoral pain: A qualitative study. Phys Ther Sport. 2022;56:1-7.
159. Vicenzino B, Maclachlan L, Rathleff MS. Taking the pain out of the patellofemoral joint: articulating a bone of contention. Br J Sports Med. 2019;53(5):268-9.
160. Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. Arthritis Rheum. 2000;43(5):995-1000.
161. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum. 2008;59(9):1207-13.
162. Hart HF, Barton CJ, Khan KM, Riel H, Crossley KM. Is body mass index associated with patellofemoral pain and patellofemoral osteoarthritis? A systematic review and meta-regression and analysis. Br J Sports Med. 2017;51(10):781-90.
163. Henriksen M, Nielsen SM, Christensen R, Kristensen LE, Bliddal H, Bartholdy C, et al. Who are likely to benefit from the Good Life with osteoArthritis in Denmark (GLAD) exercise and education program? An effect modifier analysis of a randomised controlled trial. Osteoarthritis Cartilage. 2022.
164. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley KM, et al. Does foot mobility affect the outcome in the management of patellofemoral pain with foot orthoses versus hip exercises? A randomised clinical trial. Br J Sports Med. 2020;54(23):1416-22.
165. Lankhorst NE, van Middelkoop M, van Trier YD, van Linschoten R, Koes BW, Verhaar JA, et al. Can we predict which patients with patellofemoral pain are more likely to benefit from exercise therapy? A secondary exploratory analysis of a randomized controlled trial. J Orthop Sports Phys Ther. 2015;45(3):183-9.
166. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-7.
167. van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. BMJ. 2009;339:b4074.
168. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev. 2015;1:CD010387.
169. Barton CJ, Crossley KM. Sharing decision-making between patient and clinician: the next step in evidence-based practice for patellofemoral pain? Br J Sports Med. 2016;50(14):833-4.
170. Legare F, Stacey D, Turcotte S, Cossi MJ, Kryworuchko J, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev. 2014(9):CD006732.

171. Smith BE, Moffatt F, Hendrick P, Bateman M, Rathleff MS, Selfe J, et al. The experience of living with patellofemoral pain-loss, confusion and fear-avoidance: a UK qualitative study. BMJ Open. 2018;8(1):e018624.

Appendices for PhD thesis

Appendix 1 – Clinical Study Protocol

CLINICAL STUDY PROTOCOL

COMPARATIVE EFFECTIVENESS OF THERAPEUTIC HIP AND KNEE EXERCISE FOR PATELLOFEMORAL PAIN: A PRAGMATIC RANDOMISED TRIAL

STUDY NAME

COMPETE

Study Director & Principal Investigator:	Marius Henriksen, PT, PhD
Investigator:	Rudi Hansen, PT, MSc
Investigator:	Christoffer Brushøj, MD, PhD
Investigator:	Michael Rathleff, PT, PhD
Sponsor:	Department of Physical and Occupational Therapy, Bispebjerg-Frederiksberg Hospital, Bispebjerg Bakke 23, DK-2400 NV, Denmark

1.0 STUDY COMMITTEES

Steering committee:

Name	Title & Position	Affiliation
Marius Henriksen (Chairman)	PT, PhD, Professor	Department of Physical and Occupational Therapy, Bispebjerg-Frederiksberg Hospital The Parker Institute, Bispebjerg-Frederiksberg Hospital
Rudi Hansen	PT, MSc	Department of Physical and Occupational Therapy, Bispebjerg-Frederiksberg Hospital
Christoffer Brushøj	MD, PhD	Institute of Sports Medicine Copenhagen, Bispebjerg-Frederiksberg Hospital
Michael Rathleff	PT, PhD	Research Unit for General Practice in Aalborg and Department of Clinical Medicine at Aalborg University

2.0 FLOW CHARTS

2.1 Study flow chart

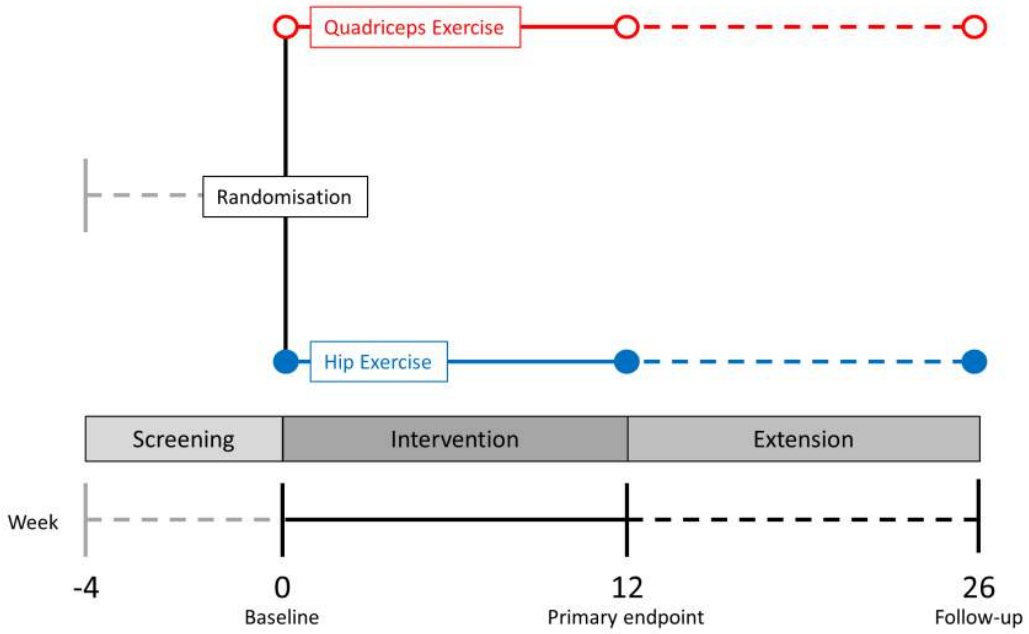


Figure 1. Graphical illustration of the study design

3.0 TABLE OF CONTENTS

4.0 TRIAL IDENTIFIER

4.1 Full title of trial

Comparative effectiveness of therapeutic hip and knee exercise for patellofemoral pain: a pragmatic randomised trial.

4.2 Short title

Comparison of exercise therapies for patellofemoral pain

4.3 Acronym

The COMPETE study

4.4 Health Research Ethics Committee Number

H-16045755 (approved December 15, 2016)

4.5 Trial registration identifier and date

www.clinicaltrials.gov: NCT03069547, registered February 22, 2017

4.6 Internal registration number

FYS-2016-004

4.7 Version number and date

Version 1.2

April 18, 2017

4.8 Revision history

Version #	Issue date	List of major changes
1	November 11, 2016	This is the authorisation version
1.1	December 13, 2016	Minor changes prompted by Health Research Ethic Committee
1.2	April 18, 2017	Trial registration identifier added Minor adjustment to exploratory outcome (section 10.6.11)

4.9 Sponsor

Department of Physical and Occupational Therapy, Copenhagen University Hospital Bispebjerg - Frederiksberg.

5.0 BACKGROUND

Patellofemoral Pain (PFP) is a common knee problem, which particularly affects adolescents and young adults (1-4). PFP is characterised by significant retropatellar and/or peripatellar pain and impairment of function and quality of daily life (5). Exercise has repeatedly been shown beneficial for pain and physical function and is unequivocally recommended as a core component of the management of PFP (6, 7). Different types of exercise (e.g. quadriceps strengthening, hip strengthening and functional/neuromuscular exercises) have been investigated. In general, these different types of exercises produce similar small to moderate beneficial effects in pain and physical function (8). However, the PFP population is very heterogeneous and "one-size-fits-all"-approaches presumably are sub-optimal because the heterogeneity is ignored. This heterogeneity probably explains the overall limited beneficial effects of exercise, and the lack of differences in direct comparisons of different exercise types.

Extensive effort has been made to identify indicators of prediction for conservative PFP management (9). Studies included in a recent systematic review on outcome predictors for conservative patellofemoral pain management included only cohort studies and no RCT's (2). Therefore the outcome predictors have a high risk of being non-specific predictors of outcome without causal relationship to a specific intervention. Furthermore, most of the outcome predictors identified requires the use of expensive and inaccessible equipment (e.g. magnetic resonance imaging) and is therefore not feasible in a clinical setting.

Based on current knowledge it is not possible to identify one exercise protocol that yields superior effects, and therefore current recommendations emphasises a multimodal intervention including exercise to strengthen the gluteal and quadriceps musculature, manual therapy and taping (7).

Traditionally, quadriceps based exercise programs have been recognised as a cornerstone in PFP rehabilitation (10). It is theorised that patella instability and/or maltracking causes the retro- and peripatellar pain (11), and that strength training of the quadriceps muscle improves stability and alignment, potentially relieving the pain.

Isolated hip abductor and external rotator strengthening has also shown promising results on pain and health status in females with PFP. The rationale for applying hip strengthening is that the patellofemoral joint may be stressed by excessive dynamic valgus and medial femoral rotation during weight bearing activities (12). Therefore strengthening the hip abductors and external rotators potentially modifies the biomechanics and may be beneficial in the rehabilitation process. An

increasing quantity of literature supports the implementation of hip strengthening exercises in the rehabilitation process (6, 13-15).

By training the hip muscles in isolation, patellofemoral stress is avoided during exercise and the concomitant pain and discomfort circumvented. This may yield a higher degree of compliance to the rehabilitation program as exercise induced pain flares presumably are fewer and less intense. There is, however, insufficient evidence to determine the best form of exercise therapy and it is unknown whether these results would apply to all people with PFP (8). It can be hypothesised that certain patient characteristics may predict outcome success of either a hip training program or a training program that focus on quadriceps training, but this remains to be shown.

The purpose of this study is to compare changes in pain and function for patients assigned to a focused "Quadriceps Exercise" protocol (QE) and a "Hip Exercise" protocol (HE) for a 12-week period. Furthermore, we seek to identify clinically feasible outcome predictors of treatment success of either protocol making it possible for clinicians to target the intervention in the future. To our knowledge no randomised controlled trial has been conducted with the aim to identify outcome predictors.

6.0 STUDY OBJECTIVES & HYPOTHESES

6.1 Principal research questions to be addressed

This study has two aims:

- To assess the comparative effectiveness of two different exercise programs (QE vs. HE) on self-reported pain and function in individuals with PFP.
- To explore candidate patient characteristics that predict differential responses to the two exercise programs (QE vs HE) on self-reported pain and physical function in individuals with PFP.

6.2 Hypothesis

According to the study aims (above) we pursue the following hypothesis:

1) QE and HE have equivalent efficacy on self-reported pain after 12 weeks of treatment in patients with PFP.

As stated in the second study aim above we aim to explore possible candidate patient characteristics that may associate with differential outcomes. As this is exploratory, the pursuit of this aim is hypothesis-free.

7.0 STUDY DESIGN

7.1 Description of the protocol

This is an investigator-initiated study that will be registered at www.ClinicalTrials.gov before initiation. The study is a randomised parallel-group equivalence trial with blinded outcome assessors aiming to compare the effectiveness of a 12-week rehabilitation program focused on quadriceps exercises (QE) with a 12 week hip exercise (HE) program in individuals with PFP and to investigate if certain clinical characteristics associate with better outcomes of one of the exercise programs. The investigators and outcome assessors will attempt to remain unaware of the group assignments. All participants will be recalled for clinical assessment visits every 4 weeks until week 12 (primary endpoint and end-of-trial) with assessment of clinical effectiveness and safety.

Six months (26 weeks) after randomisation participants who have completed the week 12 visit are invited to participate in an internet-based survey including the patient reported outcomes used in the main trial phase.

7.2 Rationale for the study design

7.2.1 Rationale for dose, duration, and observation period

Exercise programs for muscle strength have optimum effectiveness if lasting 12 weeks or more, with 3 weekly exercise sessions (16). To explore the long-term symptomatic benefits and possible associations with baseline patient characteristics, an internet-based survey is conducted 24 weeks after randomisation.

7.3 Randomisation, and reduction of bias

By using randomisation we include a deliberate element of chance into the allocation assignment of the participants in the study. Eligible participants will be randomly assigned - in permuted blocks of 4 and 6 - according to a computer-generated list of random numbers, to one of the two following groups:

- QE – quadriceps exercise program
- HE – hip exercise program

A duplicate of the randomisation list will be stored under lock at the research administration of the Department of Physical and Occupational Therapy.

Following the inclusion of a participant, delivery of informed consent and all baseline assessments, the including member of the study staff will assign a randomisation number to the participant according to the randomisation list.

7.4 Randomisation technique

The study director is responsible for preparation of the randomised group allocation list before initiation of the trial. The following randomization technique is used:

Eligible participants will be randomly assigned - in permuted blocks of 4 and 6 - according to a computer-generated list of random numbers, to one of the two groups (QE or HE).

The randomisation will be equal (1:1), meaning that 100 participants are randomly assigned to each group. A coded randomisation list will be available to the clinical staff administering the interventions.

After inclusion of a participant, collection of signed informed consent from the participant, and completed baseline measurements, the clinical staff will allocate participants according to the randomisation.

The participants are randomised using an envelope-based randomization technique. Prior to the study initiation, a folder containing 200 envelopes is created. Each envelope will contain a piece of paper with the randomisation written on it representing the group allocation (i.e. “QE”, or “HE”). The order of the envelope content matches the randomisation list. It is ensured that the envelopes are closed and opaque. The envelopes are numbered consecutively from 1-200 and placed in the folder according to the number (no. 1 in front – no. 200 at the back). The envelope numbers match the randomisation numbers. The folders are stored in a locked locker in the principal investigator’s office. Duplicates of the randomisation list and envelopes are stored under lock in the research administration at the Department of Physical and Occupational Therapy.

When a participant is ready for randomisation an envelope is drawn from the folder, starting from the front envelope. At the front of the envelope, the participant’s name, date of birth, screening number, and present date is written.

7.5 Blinding

Investigators, study coordinators, clinical staff, study staff, and other personnel directly involved in the study, will be blinded to the group allocation.

Participants and staff involved in the exercise delivery are not blinded to the group allocation. Information that could potentially unblind otherwise blinded staff will not be shared, and will be stored in facilities with limited access until the study is completed. Unblinding of blinded personnel does not preclude the related participants' continued participation in the study.

7.6 Allocation concealment

To ensure concealment of the assigned intervention, the randomisation list is stored securely in facilities with limited access. Individual allocation is kept in in numbered, opaque, and sealed envelope containing the participant's assigned intervention.

Blinding or masking are intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial, arising from the influence which the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of endpoints, the handling of withdrawals, the exclusion of data from analysis, and so on. As this study cannot be appropriately masked for the participants, who will also evaluate the treatment using Patient-Reported Outcomes, we are aware of the implied risk of performance bias - and to some extent detection bias - due to knowledge of the allocated interventions by participants and some part of the study staff during the study.

The investigators and outcome assessors will attempt to remain unaware of the group assignments, and the allocation will remain concealed to any staff associated with the study. This, however, may be difficult to ensure, because the study staff inevitably will have some interaction with the study participants.

7.7 External data monitor

As this in principle is a non-blinded trial, there is no need for an independent data safety monitoring committee.

7.8 Stopping rules

The study is terminated if the safety or health of the participants is compromised. If a possible safety issue arises, the study is suspended until the issue has been investigated and it has been established if the issue is related to any of the interventions or assessment procedures associated with the study.

8.0 PARTICIPANTS

8.1 Number of participants planned

It is anticipated that approximately 200 participants will be enrolled in this study. A participant may be enrolled in this study provided he/she has met all of the inclusion criteria and has not met any of the exclusion criteria specified below.

8.2 Recruitment

Participants will be recruited from the Institute of Sports Medicine Copenhagen (ISMC), Bispebjerg-Frederiksberg Hospital, Denmark that is a medical unit for patients with injuries in the musculoskeletal system caused by participation in sports activities.

Patients are referred to ISMC from general practitioners in all age groups. The patients are mostly amateurs, but there are also professional athletes.

All potential participants will be assessed by the Investigator (Christoffer Brushøj) or one of his delegates who are sports medical doctors employed at ISMC.

8.3 Inclusion criteria

Potential participants are eligible for the study if they meet the following criteria:

- A clinical diagnosis of PFP in at least one knee confirmed by an experienced sports medicine doctor.
- Visual analogue score rating of pain during activities of daily living during the previous week at a minimum of 3 on a 10 cm scale.
- Insidious onset of symptoms unrelated to trauma and persistent for at least 4 weeks.
- Pain in the anterior knee associated with at least 3 of the following:
  - During or after activity
  - Prolonged sitting
  - Stair ascent or descent
  - Squatting

8.4 Exclusion criteria

Potential participants are ineligible for participation if they meet one or more of the following criteria:

- Meniscal or other intra-articular injury
- Cruciate or collateral ligament laxity or tenderness
- Patellar tendon, iliotibial band, or pes anserine tenderness

- Osgood-Schlatter or Sinding-Larsen-Johansson syndrome
- History of recurrent patellar subluxation or dislocation
- History of surgery to the knee joint
- History of head injury or vestibular disorder within the last 6 months

9.0 TREATMENTS

The two exercise programs are scheduled to last 12 weeks. The programs are home-based and transferable to a clinical setting in which it is not possible to monitor the patients on a daily or weekly basis. The home-based exercise sessions are scheduled for 3 times per week with a monthly follow-up at a physiotherapist (clinical visit) with adjustment of technique and intensity. At least 48 hours of restitution is recommended between training sessions.

The training sessions are scheduled to last approximately 20-30 minutes including a 5 minutes warm up.

At the first clinical visit (after allocation), the participants are introduced to and instructed in the exercises by an experienced physiotherapist according to the allocation. The aim of the instruction is that the participant is able to do one set of each exercise with satisfying quality.

Focus of the exercises is on the quality of the performance, possibly at the expense of intensity. It is of utmost importance for gaining as much of the exercises as possible to do the exercises correctly and this issue is emphasized during the instruction session.

Each program consists of three resistance training exercises. Detailed descriptions of the two programs are given in Appendix A, and briefly summarised below.

9.1 Hip exercise program

The exercises used in the HE-group program (Appendix A) have been tested to be effective in recruiting the gluteal muscles maximally, as validated by electromyographical measurements (17-19), and are widely used in clinical practice. The exercises have been chosen due to their effectiveness and because they put minimal strain on the knee and patellofemoral joints. The exercises are easy to conduct and are easily progressed.
Since some aspects of the hip exercises involve weight bearing, several other muscles are recruited when performing the exercise, including the quadriceps. Thus, while the focus and the external load

is on the hip muscles, a possible parallel training of the quadriceps (and other synergistic) muscles cannot be completely ruled out.
However, as this study is pragmatic it is we believe that the importance of assessing and comparing exercise programs that are feasible and used in daily practice outweighs the importance of a complete avoidance of training overlap between the two exercise protocols.

9.2 Quadriceps exercise program

The exercises used in the QE-group program (Appendix A) are also widely used in clinical practice. The exercises are tested to be effective in recruiting the vastus medialis obliquus muscle (17), which is has been proposed to play a role in the aetiology of PFP if dysfunctional (11). Since two of the exercises are multiple-joint exercises, several muscles are recruited when performing the movement. In the weight bearing squat and lunge, the quadriceps muscles are activated in concomitance with primarily the hamstrings and several gluteal muscles. Thus, while the focus and the external load is on the quadriceps muscle, a possible parallel training of the hip (and other synergistic) muscles cannot be completely ruled out.
However, as this study is pragmatic it is we believe that the importance of assessing and comparing exercise programs that are feasible and used in daily practice outweighs the importance of a complete avoidance of training overlap between the two exercise protocols.

9.3 General exercise considerations

The overall principals of current resistance training program are based on guidelines from the American College of Sports Medicine (16). The aim of the exercises are to gain muscular strength and, through numerous neuromuscular mechanisms, to enable greater force generation. Muscular strength is defined as the ability of a muscle or muscle group to exert a maximal external force. A 2-10 % increase in load is recommended when the participant is able to perform two repetitions more than the desired number (16) with satisfying quality.

The exercises include both concentric and eccentric muscle action. Participants are instructed to perform the exercises in a moderate velocity, i.e. 1-2 seconds in the concentric movement and 1-2 seconds in the eccentric movement (16). Recommended load is 60-70% 1RM for novice to intermediate, i.e. 8-12 repetitions (repetition maximum; RM), recommended volume is three sets; recommended rest period is 1-2 minutes between the sets.

9.4 Information leaflet

All participants – irrespective of group allocation – will receive the information leaflet "*Managing my patellofemoral pain*" (Danish title: "*Håndtering af mine forreste knæsmerter*". The leaflet is targeted patients and written in an easily read language. It contains general information causes and management of PFP. The leaflet is appended this protocol in Appendix I (Danish version).

9.5 Concurrent treatments

Other exercise programs/regimes than the one the participants are allocated to may not be initiated during the main trial phase (week 1-12). Other non-pharmacological treatments are allowed. The usage of such other treatments/therapies will be recorded on the case report form (CRF).

Habitual use of pharmacological therapies is allowed and will be recorded on the CRF. Any new pharmacological therapies, or changes in ongoing therapies, will be recorded on the CRF.

10.0 OUTCOME ASSESSMENTS

Unless otherwise specified, all outcomes relate to the participants target knee, defined by the participants at inclusion as the most symptomatic knee.

10.1 Primary outcome

Change from baseline in the KUJALA scoring questionnaire at week 12

10.2 Secondary outcomes

Change from baseline at week 26 in

- The KUJALA scoring questionnaire

Change from baseline at week 12 and 26 in

- The 5 subscales of the KOOS questionnaire (Pain, Symptoms, Function, Sports/Recreation, Quality of Life)
- Isometric muscle strength of hip abductors, hip external rotators, hip extensors, and quadriceps
- Dynamic Assessment of Pain
- Pain Self-Efficacy Questionnaire
- EuroQoL EQ-5D-3L Questionnaire
- Transition Questionnaire of global perceived effect on overall health, pain, and function

10.3 Candidate baseline characteristics

- Baseline values from primary and secondary outcomes
- Demographic variables (age, sex, body mass, BMI, education, social status etc.)
- Previous treatments received
- Presence of bilateral knee pain
- Presence of pain in other lower extremity joints (feet, ankles, hips)
- Presence of low back pain
- PainDetect questionnaire
- Exercise Self-Efficacy Questionnaire
- Pain Catastrophizing Scale
- Hyper mobility
- Knee joint alignment during a forward lunge
- Knee joint alignment during a single-leg mini-squat
- Physiotherapist assessed prognosis for the participant
    - Based on group allocation
- Midfoot mobility

10.4 Follow-up internet based survey

- KUJALA score questionnaire
- KOOS questionnaire
- Pain Self-Efficacy Questionnaire
- EuroQoL EQ-5D-5L Questionnaire
- Transition Questionnaire of global perceived effect on overall health, pain, and function

10.5 Descriptions of the outcomes & Follow-up internet survey

10.5.1 The KUJALA score questionnaire

The Kujala Patellofemoral Scale - sometimes called the anterior knee pain scale - is disease specific validated disability scale ranging from 0 (complete disability) to 100 (fully functional) (20). It is a 13-item self-report questionnaire that documents response to 6 activities (walking, running, jumping, climbing stairs, squatting, and sitting for prolonged periods with knees bent), as well as symptoms such as limp, inability to weight bear, swelling, abnormal patellar movement, muscle atrophy, and limitations in knee flexion. An exemplar version of the questionnaire is given in Appendix B.

10.5.2 Knee injury and Osteoarthritis Outcome Score (KOOS) questionnaire

The Knee injury and Osteoarthritis Outcome Score (KOOS) is used to assess patient-reported knee-related symptoms (14). KOOS is a patient-reported outcome measurement instrument developed to assess the patient's opinion about their knee and associated problems. KOOS is user-friendly and takes about 10 minutes to complete. KOOS evaluates consequences of osteoarthritis (OA). KOOS comprises of 42 items in 5 separately scored subdomains: KOOS Pain (9 items), KOOS Symptoms (7 items), Function in daily living (KOOS Function; 17 items), Function in Sport and Recreation (5 items), and knee-related Quality of Life (4 items). The previous week is the time period considered when answering the questions, and the questions relate to one knee (the target knee in this trial). Standardized answer options are given (5 Likert boxes), and each question is assigned a score from 0 to 4. A normalized score (100 indicating best score and 0 indicating worst score) is calculated for each subdomain. A total score has not been validated and is not recommended. An exemplar version of the questionnaire is given in Appendix B.

10.5.3 Isometric muscle strength (quadriceps and hamstrings)

Isometric muscle strength of hip abductors, hip external rotators, quadriceps, and hamstrings is performed by using a handheld dynamometer. Measurement variation has previously been assessed and found to be less than 5% when assessing hip abduction, hip external rotation and knee extension, and less than 10% when assessing knee flexion (21-23). The muscle strength tests are conducted following the a previously published testing protocol (21).

10.5.4 Dynamic Assessment of Pain

The Functional Weight Bearing Pain Test is a simple performance test with an integrated pain score, designed to provide useful information for monitoring treatment progress and evaluating treatment effects in clinical physiotherapy practice (24). The patient is asked to perform as many squatting movements (both legs) as possible within 30 seconds. The knees should reach approximately 90 degrees of flexion and full extension for each squat. This is supervised by the rater.

There outcome of the test is the knee pain during the test on a 0-10 Numeric Rating Scale (NRS) rated immediately after the test.

The test takes about 1 minute to perform including instructions and does not require any equipment besides a stopwatch/watch. The result is a direct measure of the patient's ability to perform a repeated movement within a short timeframe and for the degree of pain during a weight bearing movement, which reflects the limitations of daily activities due to PFP.

10.5.5 Pain Self-Efficacy Questionnaire

The pain self-efficacy questionnaire is a 10-item questionnaire developed to assess the confidence people with pain have in performing activities while in pain (25). It is applicable to all persisting pain presentations, and covers a range of functions. Confidence in performing activities are rated on a 7-point (0-6) Likert scale with 0 representing not at all confident and 6 representing completely confident. A total score is calculated by summing the answers producing a score between 0 and 60. Higher scores reflect stronger self-efficacy beliefs. An exemplar version of the questionnaire is given in Appendix B.

10.5.6 EuroQoL EQ5D Questionnaire

EQ5D is a standardised patient-reported instrument for use as a measure of health outcome and quality of life. An exemplar version of the questionnaire is given in Appendix B.

10.5.7 Transition Questionnaire of global perceived change in overall health, pain, and function

In transition ratings the participants are asked at follow-up to compare their current state with the state at baseline. This approach has limitations regarding recall bias and influence of numerous known and unknown parameters. However, a combination of changes on current state ratings (KUJALA or KOOS) and a transition questionnaire (TRANSQ) may enhance the interpretation of the results of the study.

We have designed a transition questionnaire on which the participants initially answers if their current state is “unchanged, worse” or “better” compared to the baseline visit. An “unchanged” equals a transition score of 0. If the participant answers “worse”, he/she is asked to rate the degree of worsening on a 7 point Likert scale, and the corresponding scores range from -1 to -7. Correspondingly, if a participant answers “better”, he/she is asked to rate the degree of improvement on a 7 point Likert scale, and the corresponding scores range from 1 to 7. Thus the Transition score range from -7 (worsening) to 7 (improvement), with the mid-point – 0 – representing no change. The transition scale is used to assess overall knee related health status. An exemplar version of the questionnaire is given in Appendix B.

10.6 Descriptions of the candidate baseline characteristics

The candidate baseline characteristics encompass self-reported information as well as clinical observations and tests. The self-reported information will be included in a baseline questionnaire that is appended (Appendix B).

10.6.1 Previous treatments

The participants are asked to report any previous treatments they have received for their PFP (target knee only). The following options are pre-defined: Exercise therapy, oral analgesics, injection(s), taping, bracing, orthoses, and manual therapy. Further ample space for free text is provided.

10.6.2 Low back pain

Self-reported presence and frequency of low back pain. The participants are asked to estimate the frequency of low back pain during that last 3 months. Predefined frequency options are available:

- Almost daily
- Several times during a week
- Weekly
- Monthly
- Rarely

We will a priori define a dichotomisation of the scores as “frequent low back pain” (“Almost daily” and “several times during a week”) and “infrequent low back pain” (“Weekly”, “Monthly”, and “Rarely”). We will make sensitivity tests of this a priori cut-off by changing the cut-off value by +/- frequency option.

10.6.3 Presence of bilateral knee pain

Self-reported presence and frequency of pain in the contralateral knee (not target knee). The participants are asked to estimate the frequency of pain in contralateral knee pain during that last 3 months. Predefined frequency options are available:

- Almost daily
- Several times during a week
- Weekly
- Monthly
- Rarely

We will a priori define a dichotomisation of the scores as “frequent bilateral knee pain” (“Almost daily” and “several times during a week”) and “infrequent bilateral knee pain” (“Weekly”, “Monthly”, and “Rarely”). We will make sensitivity tests of this a priori cut-off by changing the cut-off value by +/- frequency option.

10.6.4 Presence of pain in other lower extremity joints (feet, ankles, hips)

Self-reported presence of pain in the feet, ankles, and/or hips lasting more than 3 months, recorded as "Yes" (pain present) or "No" (No pain) for feet, ankles and hips, respectively. The participants are asked to estimate the frequency of pain in other lower extremity joints (feet, ankles, and hips) during that last 3 months. Predefined frequency options are available:

- Almost daily
- Several times during a week
- Weekly
- Monthly
- Rarely

We will a priori define a dichotomisation of the scores as "frequent [foot/ankle/hip] pain" ("Almost daily" and "several times during a week") and "infrequent [foot/ankle/hip] pain" ("Weekly", "Monthly", and "Rarely"). We will make sensitivity tests of this a priori cut-off by changing the cut-off value by +/- frequency option.

10.6.5 PainDetect questionnaire

The painDETECT questionnaire (PDQ) is a patient reported questionnaire developed and validated to assess presence of signs of neuropathic pain in patients with chronic low back pain (26). It comprises questions regarding pain intensity, course of pain, subjective experience of a radiating quality of the pain, and the presence and perceived severity of seven somatosensory symptoms of neuropathic pain. A validated algorithm is used to calculate a total score ranging from 0 to 38. A score ≤12 indicates that the presence of neuropathic pain is unlikely, a score ≥19 indicates that a neuropathic pain component is likely to be present, while a score of 13–18 points towards unclear screening conclusion.

10.6.6 Exercise Self-Efficacy Questionnaire

Exercise self-efficacy in relation to the two different exercise programs are assessed by asking the participants to rate their confidence in performing the two different exercise programs on a 7-point (0-6) Likert scale with 0 representing "*Not at all confident*" and 6 representing "*Completely confident*".

10.6.7 Pain Catastrophizing Scale

The Pain Catastrophizing Scale (PCS) (27) will used as a measure of pain-related catastrophic thinking. The PCS instructs participants to reflect upon past painful experiences, and to indicate the

frequency with which 13 pre-specified thoughts or feelings occur while they are experiencing pain. The frequency is scored on a 5-point scale ranging from 0 (not at all) to 4 (all the time). The PCS measures 3 distinct components: rumination, magnification, and helplessness.

10.6.8 Hypermobility

Hypermobility is assessed by the Beighton Score (28) applying the revised criteria for the diagnosis of benign joint hypermobility syndrome (29, 30). The Beighton score ranges from 0-9. In this study we define hypermobility as a score of 4 or more (29). Details are given in Appendix C.

10.6.9 Knee joint alignment during a forward lunge movement

Knee joint alignment is assessed by clinically observation of the participant while he/she performs a forward lunge movement. The observation is made by a trained physiotherapist that classifies each participant's knee as "*varus*", "*neutral*", or "*valgus*".
Detailed description of the assessment protocol is given in Appendix D.

10.6.10 Knee joint alignment during a single-leg mini-squat movement

Knee joint alignment is assessed by clinically observation of the participant while he/she performs a single-leg mini-squat movement. The observation is made by a trained physiotherapist that classifies each participant's knee as "*varus*", "*neutral*", or "*valgus*".
Detailed description of the assessment protocol is given in Appendix D.

10.6.11 Physiotherapist estimated prognosis for the participant

A trained physiotherapist's estimate of each patient's potential for a successful outcome based on professional appraisal. The physiotherapists will be asked to appraise all component parts of their evaluation in their prediction of prognosis of each patient.

The prognosis will be done based on group allocation (i.e. after allocation) judging the participant's prognosis based on the allocated treatment

The therapists are instructed to score each patient on a continuum of 1 (suggesting a very poor projected outcome) to 10 (suggesting an excellent projected outcome). The therapist must score each patient following their complete encounter with the patient. This may include the physiotherapist's assessment of the patient's resources (personal, material, social, etc.), personality, medical history, comorbidities, surgery reports, physical examination(s), physiotherapeutic assessments, and more.
The prognosis assessment forms are appended in Appendix E.

10.6.12 Midfoot mobility

Midfoot mobility is calculated as the change in midfoot width from non-weight bearing to weight bearing. Midfoot width is measured at 50% of total foot length using a caliper during non-weight bearing and during standing with equal weight on each foot. The mobility is recorded as the difference between non-weight bearing and weight bearing measurements.

11.0 DISCONTINUATION

11.1 Participant withdrawal

A participant may withdraw from the study at any time without this impacting on any future investigations and/or treatments at the site, by the Investigators in this study or by other staff associated with the study.

If a participant withdraws from the study, the procedures outlined for the FU Visit is sought to be completed within 2 weeks of the last rehabilitation session, and preferably prior to the initiation of another therapy. However, these procedures should not interfere with the initiation of any new treatments or therapeutic modalities that the investigator feels are necessary to treat the participant's condition.

It is important to avoid any lost to follow-up participants for the efficacy assessment and meaningful analysis of the study.

11.2 Individual participant discontinuation

Subjects may discontinue the study at any time after receiving study intervention. The reason for discontinuation will be recorded on the CRF.

Criteria for individual subject discontinuation from treatment include, but are not limited to, the following:

- Any medical condition that, in the opinion of the investigator, may jeopardise the participant's safety if he or she continues in the study
- Noncompliance with study schedule or procedures

11.3 Discontinuation of Entire Study

The Principal Investigator has the right to terminate this study at any time. Reasons may include the following, but are not restricted to:

- The incidence of events in this or other studies that indicate a potential health hazard to subjects
- Unsatisfactory subject enrolment

12.0 STUDY PROCEDURES

12.1 Time plan

The study will be initiated once necessary funding has been acquired (pending) and this protocol has been approved by relevant authorities (pending).
Enrolment of participants is expected at a rate of 4 per week, resulting in an enrolment phase lasting approximately 24 months. Last participant's last visit (LPLV) is scheduled after additional 26 weeks.

12.2 Visit windows

The visit windows are as follows:

- Baseline (BL) measurements will be taken no more than 28 days before randomisation.
- Outcome assessments can be taken within +/-7 days for the scheduled visits at week 13.
- The 26 week assessment can be taken within +/-14 days for the scheduled visit.

12.3 Screening Period

At the first (screening) visit, participants will receive a full explanation of the study design and study procedures, provide written informed consent and undergo the screening procedures. Patient who meet all inclusion criteria and who do not have exclusions will be scheduled for BL visit up to 28 days prior to randomisation.

12.4 Baseline and follow-up visits

Eligible participants scheduled for rehabilitation in the department of physical and occupational therapy, and who have been informed orally and in writing about the study, and provided informed consent, are included in the study.

Upon inclusion, the participant designates a target knee (defined as the most symptomatic knee), that will be the target for most of the subsequent assessments (unless otherwise specified). Subsequently, baseline questionnaires and measurements are performed, and after completion of these, the participants are randomised and allocated. All outcome measurements are repeated at the 12-week follow-up visit.

12.5 End-of trial visit (week 26)

The study is ended when the last participant has completed the final outcome assessments. For individual participants study participation is considered ended after his/her final outcome assessments. If a participant presents an occurred abnormality at the final examinations or a lingering side effect of the intervention, the participant is followed until normalisation of the abnormality or the condition can be described as stable or chronic.

13.0 SUB-STUDIES

Sub-studies of high scientific merit may be conducted based on the recommendation of the Steering Committee. These studies will merely be comprised of additional observations and measurements and will not interfere with the treatment. Sub-studies will be formulated in separate protocols and will depend on separate funding. The reporting of the sub-studies will be separated from the main trial.

Currently, one sub-study is planned on a sub-group of participants, with details provided in Appendix D. Participation in the sub-study is not considered a requirement for participation in the main study. Nor will any violation of the sub-study protocol affect participation, care, or attention given from the parent study.

This sub-study's procedures are included in the main study written patient information and informed consent. The oral information will be delivered together with that of the parent study.

14.0 SAFETY

It is not expected that study participation will be associated with risks or complications. The applied interventions are both evidence based and will delivered by staff with relevant qualifications, education and certification. Thus, according to the current good clinical practice (GCP) standard, passive surveillance of harms will be assessed: The recorded adverse events are those that the study participants spontaneously report on their own initiative.

15.0 DETERMINATION OF SAMPLE SIZE AND STATISTICAL ANALYSIS PLAN

15.1 Study participants description

15.1.1 Disposition of participants

The number of randomized patients will be summarized as total and by site using counts and percentages. The number of patients either completing or permanently discontinuing the study will be summarized using counts and percentages.

15.2 Study Population definitions

15.2.1 Intention to treat population (ITT)

The ITT population consist of all randomized patients irrespective of whether the patient actually received study intervention or the patient's compliance with the study protocol, in the treatment group to which the participant was assigned at randomisation. A patient will be considered randomized as soon as a treatment is assigned according to the allocation sequence.

The mITT (modified intention-to-treat) population consists of participants in the ITT population *with* a valid baseline observation of the variable to be analysed.

15.2.2 As-observed population (AO)

The AO population consists of participants who has the outcome of interest assessed at baseline and week 12 (i.e. no imputation of missing data will be done).

15.2.3 Per-protocol population (PP)

The PP population is defined as the AO population participants that adhere to this protocol, defined by fulfilment of all of the following criteria:

1. Is included in the AO population, AND
2. Have performed at least 2/3 of the scheduled home-exercise sessions, AND
3. Complies with the rules for concomitant treatments as described above.

As sensitivity analyses we might adjust the second criteria to e.g. 75%, 80%, or other percentages of performance of the scheduled sessions.

15.3 Analysis of the conduct of the study

The disposition of participants will be summarised and analysed for both PP and ITT populations. The number of participants who violate key eligibility criteria will be summarised. For subjects who discontinue the study early, reasons for discontinuation will be summarised and listed according to treatment allocation.

15.4 Analysis of treatment group comparability

Analyses of treatment group comparability will be performed for PP, AO, and ITT populations.
Demographic variables and baseline characteristics such as age, sex, race, weight and height, and prognostic factors such as baseline pain measurement and radiographic disease will be summarised

by treatment group. Means, standard deviations, medians, and ranges will be used for continuous variables. Counts and proportions will be used for categorical variables.

15.5 Analysis of equivalence

All equivalence analyses (study aim 1) are carried out on the per-protocol (PP) population. Changes from baseline in outcomes are compared between groups using repeated measures mixed linear models with group (QE vs HE) and week (12 and 26) as fixed factors, and participant as random factor. The models will be adjusted for the outcome baseline value. The study hypotheses of equivalence can only be accepted if a statistically non-significant difference between groups (accepting the null hypothesis) in the primary outcome (change in KUJALA) is found and the 95% confidence intervals respect the equivalence margins of +/- 8 KUJALA points. All statistical tests will be 2-tailed at an alpha = 0.05 with estimates presented with 95% confidence intervals.

15.6 Analysis of effectiveness according to patient subgroups

The analyses of effectiveness are based on assumptions that treatment effects vary according to an interaction between group allocation and certain patient characteristics that can be binary defined as present/absent.

The analyses will be done on the AO population as defined above.

The AO population will be split in two; an exploration cohort and a replication cohort. The split will be done randomly on a 1:1 basis (equal group sizes). A randomisation scheme will be prepared, and the patients will be randomly assigned in blocks of 4-6. The randomisation will be stratified for group allocation and for the presence/absence of the variable of interest. This may result in multiple randomisations. Randomisation will occur before any analytical steps have been taken.

The split is made in order to replicate our results immediately (i.e. without running another study). We will run all analyses (see below) in the exploration cohort and assess if the findings can be replicated in the replication cohort.

The analyses will focus on the interaction between presence/absence of a certain characteristic and group allocation (QE vs. HE) at each time point (week 12 and 26). Changes from baseline in outcomes will be analysed using repeated measures mixed linear models with group (QU vs HE), week (12 and 26), and candidate baseline predictor characteristic (present/absent) as fixed factors, and participant as random factor. The models will be adjusted for the outcome baseline value. We will focus on the

triple interaction part of the model (group×week×predictor) that will be broken down by pair-wise comparisons of group and predictors at each week. All statistical tests will be 2-tailed at an alpha = 0.05 with estimates presented with 95% confidence intervals.

15.7 Sensitivity analyses

If sensitivity analyses are deemed necessary, these will be repeated on the ITT and mITT populations, as defined above.

15.8 Missing data

For any analyses on the ITT and mITT populations, we will replace data using baseline-observation carried forward (BOCF; non-responder imputation).

15.9 Sample size and statistical power considerations

The sample size is determined to be 200 participants equally allocated (1:1). The sample size has been reached based on the following calculation:

Sample size calculation relating to study aim 1:

The sample size is calculated to test the equivalence of the QE and HE programs in the assessment of change in the KUJALA pain subscale. With 77 participants per group, the study will have 90% power assuming the expected group difference in mean changes from baseline is 0, the common standard deviation is 15 (0-100 scale), with a delta (equivalence margin) of 8 units (0-100 scale) corresponding to the suggested minimum clinically relevant difference, and a significance level of 5%. With an expected drop-out during the study we will randomise and allocate 200 participants (100 to each group); analysing the PP population.

Sample size calculation relating to study aim 2:

We have no presumptions about the multiple group*predictor interactions that we pre-specify to explore. However, we believe the conservatively set sample size (above) has sufficient power to reliably detect candidate phenotypical characteristics that may associate with differential treatment response.

16.0 HEALTH RESEARCH ETHICS

16.1 General considerations

Prior to screening, all potential trial participants are informed, both orally and in writing, about the trial's purpose, process and potential risks, costs and benefits of participation. In addition, the leaflet

'*Rettigheder som forsøgsperson i et sundhedsvidenskabeligt forskningsprojekt*' is handed out. All participants are informed of their rights to withdraw from the study at any time without this impacting on any future investigations and/or treatments at Copenhagen University Hospital at Bispebjerg and Frederiksberg, or by some of the members of the study group. After the information is delivered, read and understood, voluntary informed consent is given by the participant by signing a consent form before trial participation can take place.

16.2 Oral information

When a potential participant is identified, an appointment for an information interview is made. It will be stressed that the investigator is asking the participant to consider participation in the trial, and that the potential trial participant has the right to bring a companion to information interview. The written information material will be sent by e-mail or regular post to the potential trial participant so that he or she has at least 24 hours to reading the material before the information interview.

The oral information is based on the written information and will be given in a language easily understood without technical or value-laden terms. The information will be given in a considerate way that is tailored to each potential trial participants. The aim is that the conversation takes place without interference. It is the responsibility of the interviewer to ensure that the potential trial participant has understood the information. The information interview is performed by the investigator and in his absence by a designated delegate. Guidelines for the oral information are given in Appendix F (document in Danish).

16.3 Written information

A written information material has been prepared and is attached this protocol as Appendix G (document in Danish).

16.4 Informed consent

Consent to participation in the trial is given on the basis of the written and oral information.
An informed consent form (ICF; Appendix H; document in Danish) has been prepared. The form must be signed and dated by the participants prior to participation in the trial. A copy of the form is provided to the participants. The investigator and his designated delegates can receive the signed consent form. Prior to consent, it must be ensured that a potential participant has been given sufficient time to consider his or her participation.

The case report forms (CRFs) will document for each subject that informed consent was obtained prior to participation in the study. The signed ICF must remain in each participant's CRF and must be available for verification at any time.

16.5 Research ethics - the interventions

Exercise therapy is recommended for PFP, and therefore all participants are provided with evidence based treatments irrespective of group allocation. There are no known risks or predictable harms associated with the interventions that at worst are considered harmless. The interventions are considered to be justified from a health research ethics perspective.

16.6 Research ethics - the outcome measures

None of the planned assessments are invasive or associated with known risks or harms. The knowledge gained by this trial is commensurate with the efforts and difficulties associated with participation. The outcomes and assessments are considered to be justified from a health research ethics perspective.

16.7 Research ethics approval

This protocol, the ICF, written patient information, any anticipated advertising materials, and relevant supporting information must be submitted to the ethical committee, by the Sponsor, prior to study initiation. The study will be conducted in accordance with Danish law, the Helsinki declaration, and local research ethics committee requirements.

The Sponsor is responsible for keeping the health research ethics committee informed of amendments or changes to the protocol, and the progress of the study, as appropriate.

17.0 SOURCE DOCUMENTS AND CASE REPORT FORM COMPLETION

17.1 Source Documents

Source documents are defined as original documents, data and records. This may include questionnaires, hospital records, clinical and office charts, laboratory data/information, participants' diaries or evaluation checklists, health professionals' records or charts, pharmacy dispensing and other records, recorded data from automated instruments, MRI images, and/or x-rays. Data collected during this study must be recorded on the appropriate source documents.

The investigator(s)/institution(s) will permit study-related monitoring, audits and regulatory inspection(s), providing direct access to source data documents.

17.2 Case Report Forms

The study will use electronic case report forms (CRF) using an on-line web-based clinical trial management application (EasyTrial). EasyTrial allows individual patients to supply data from home and clinical and objective data is entered by staff at the clinical centres. Thus maximal data completeness is ensured.

The application meets all regulatory standards, and allows management of all activities related to clinical trials that ensures optimal resource use and safety according to good clinical practice and data protection legislation.

18.0 REGULATORY STANDARDS

18.1 Administrative structure

The Department of Physical and Occupational Therapy is responsible for the entire conduct of the study.

18.2 Compliance with Laws and Regulations

The Department of Physical and Occupational Therapy will conduct this clinical research study under related legislation and regulations, including the Act on Processing of Personal Data. The Department of Physical and Occupational Therapy will comply with International Conference on Harmonisation (ICH) E6 Good Clinical Practice (GCP): Consolidated Guidance, and other national laws and regulations, as applicable.

18.3 Notification to the Danish data protection agency

Because the study is initiated from a hospital department, it is regarded as "public" in accordance with the Data Protection Agency (DPA) guidance. Therefore the notification of the study to the DPA is handled by the public authority to which the hospital department belongs, in this case the Capital Region of Denmark. Thus, the study is notified to the local DPA representative of the Capital Region of Denmark, who is responsible for the further notification to the DPA.

18.4 Study monitoring requirements

As trial primarily aims at addressing relief of symptoms, and the trial population is not at elevated risk of more severe outcomes, no external monitoring is required. However, one or more authorised representatives of The Department of Physical and Occupational Therapy may periodically inspect study data and CRFs in accordance with GCP.

18.5 Study completion

The following data and materials are required before a study can be considered complete or terminated, including, but not limited to: Clinical data, laboratory and other test results, and any special test results from screening thru the end of study for all randomised subjects. The Principal Investigator and study personnel are responsible for supplying the documentation specified.

18.6 Subject confidentiality

Subject medical information obtained by this study is confidential, and disclosure to third parties other than those noted below is prohibited.

With the subject's permission, medical information may be shared with his or her personal physician or with other medical personnel responsible for the subject's welfare.

When the data from this study are published, the presentation format will not include names, recognisable photos, personal information, or other data which compromises the anonymity of participants.

18.7 Retention of records

Danish regulations require that the records and documents pertaining to this study must be retained by the Investigator for 5 years after completion of the study. Records to be retained include, but are not limited to, CRFs, consent forms, source documentation, test results, medication inventory records, and regulatory documents.

18.8 Quality assurance

All data will be entered into a study database for analysis and reporting. Any data captured electronically will be stored electronically in a separate database according to standard procedures at The Department of Physical and Occupational Therapy. Upon completion of data entry, the databases will be checked to ensure acceptable accuracy and completeness. System backups and record retention for the study data will be consistent with The Department of Physical and Occupational Therapy standard procedures.

Individuals involved in study evaluations will be trained to perform the efficacy evaluations and activity measurements described in the protocol.

18.9 Financing and insurance information

This study is initiated by Physiotherapist Rudi Hansen, MSc. and Professor Marius Henriksen PhD, who is the principal investigator of this study.

The study has received no funding. All future sources of support (including technical and financial support) provided for this study will be disclosed in the written information material and in publication of the study results. Any future financial support will be paid to a dedicated account administered by the Department of Physical and Occupational Therapy at Bispebjerg-Frederiksberg Hospital under revision from the Capitol Region of Copenhagen (Region Hovedstaden). This information will be included in the written information material if financial support should be obtained. Further, the Health Research Ethics Committee will be informed about any future financial support.

The participants are insured by the Danish Patient Insurance Association. Financing and insurance issues are addressed in the written information.

19.0 PUBLICATION

All outcomes will be published, i.e. both positive, negative and inconclusive results.

Development of the core publication will be coordinated by the executive committee, whose membership includes investigators who provided significant input into study design, implementation, conduct and interpretation. A named author approach will be utilized with Rudi Hansen as first author, Michael Rathleff as secondary author, Christoffer Brushøj as third author, and Marius Henriksen as last author. Other authors may include other key study personnel (to be agreed upon by the steering committee) who has contributed significantly to the implementation and conduct of the study and non-site personnel who contribute substantially to the design, interpretation or analysis of the study and fulfil the requirements for authorship as recommended by the international committee of medical journal editors (ICMJE).

Activities that alone (without other contributions) do not qualify a contributor for authorship include, but are not limited to: acquisition of funding; general supervision of a research group or general administrative support; and writing assistance, technical editing, language editing, and proofreading. Those whose contributions do not justify authorship may be acknowledged individually or together as a group under a single heading (e.g. "Clinical Investigators" or "Participating Investigators"), and their contributions will be specified (e.g., "served as scientific advisors", "critically reviewed the study proposal", "collected data", "provided and cared for study patients", "delivered the interventions",

"participated in writing or technical editing of the manuscript"). Written permission to be acknowledged from all acknowledged individuals will be collected prior to submission of a manuscript for publication.

20.0 REFERENCE LIST

1. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain. Am J Sports Med. 2016;44(5):1165-71.
2. Rathleff CR, Baird WN, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Hip and knee strength is not affected in 12-16 year old adolescents with patellofemoral pain--a cross-sectional population-based study. PLoS One. 2013;8(11):e79153.
3. Roush JR, Curtis Bay R. Prevalence of anterior knee pain in 18-35 year-old females. Int J Sports Phys Ther. 2012;7(4):396-401.
4. Biber R, Gregory A. Overuse injuries in youth sports: is there such a thing as too much sports? Pediatr Ann. 2010;39(5):286-92.
5. Rathleff CR, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Half of 12-15-year-olds with knee pain still have pain after one year. Dan Med J. 2013;60(11):A4725.
6. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med. 2016.
7. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'Best Practice Guide to Conservative Management of Patellofemoral Pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med. 2015;49(14):923-34.
8. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev. 2015;1:CD010387.
9. Lack S, Barton C, Vicenzino B, Morrissey D. Outcome predictors for conservative patellofemoral pain management: a systematic review and meta-analysis. Sports Med. 2014;44(12):1703-16.
10. Kooiker L, Van De Port IG, Weir A, Moen MH. Effects of physical therapist-guided quadriceps-strengthening exercises for the treatment of patellofemoral pain syndrome: a systematic review. J Orthop Sports Phys Ther. 2014;44(6):391-B1.
11. Petersen W, Ellermann A, Gosele-Koppenburg A, Best R, Rembitzki IV, Bruggemann GP, et al. Patellofemoral pain syndrome. Knee Surg Sports Traumatol Arthrosc. 2014;22(10):2264-74.
12. Powers CM. The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. J Orthop Sports Phys Ther. 2010;40(2):42-51.
13. Baldon Rde M, Serrao FV, Scattone Silva R, Piva SR. Effects of functional stabilization training on pain, function, and lower extremity biomechanics in women with patellofemoral pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2014;44(4):240-A8.
14. Fukuda TY, Melo WP, Zaffalon BM, Rossetto FM, Magalhaes E, Bryk FF, et al. Hip posterolateral musculature strengthening in sedentary women with patellofemoral pain syndrome: a randomized controlled clinical trial with 1-year follow-up. J Orthop Sports Phys Ther. 2012;42(10):823-30.
15. Dolak KL, Silkman C, Medina McKeon J, Hosey RG, Lattermann C, Uhl TL. Hip strengthening prior to functional exercises reduces pain sooner than quadriceps strengthening in females with patellofemoral pain syndrome: a randomized clinical trial. J Orthop Sports Phys Ther. 2011;41(8):560-70.

16. American College of Sports M. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.
17. Ekstrom RA, Donatelli RA, Carp KC. Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises. J Orthop Sports Phys Ther. 2007;37(12):754-62.
18. Distefano LJ, Blackburn JT, Marshall SW, Padua DA. Gluteal muscle activation during common therapeutic exercises. J Orthop Sports Phys Ther. 2009;39(7):532-40.
19. Boren K, Conrey C, Le Coguic J, Paprocki L, Voight M, Robinson TK. Electromyographic analysis of gluteus medius and gluteus maximus during rehabilitation exercises. Int J Sports Phys Ther. 2011;6(3):206-23.
20. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy. 1993;9(2):159-63.
21. Thorborg K, Bandholm T, Holmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. Knee Surg Sports Traumatol Arthrosc. 2013;21(3):550-5.
22. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. Scand J Med Sci Sports. 2010;20(3):493-501.
23. Hansen EM, McCartney CN, Sweeney RS, Palimenio MR, Grindstaff TL. Hand-held Dynamometer Positioning Impacts Discomfort During Quadriceps Strength Testing: A Validity and Reliability Study. Int J Sports Phys Ther. 2015;10(1):62-8.
24. Klokker L, Christensen R, Osborne R, Ginnerup E, Waehrens EE, Bliddal H, et al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: a reliability and agreement study. Qual Life Res. 2015;24(12):2985-92.
25. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain. 2007;11(2):153-63.
26. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22(10):1911-20.
27. Kjogx H, Zachariae R, Pfeiffer-Jensen M, Kasch H, Svensson P, Jensen TS, et al. Pain frequency moderates the relationship between pain catastrophizing and pain. Front Psychol. 2014;5:1421.
28. Beighton P, Solomon L, Soskolne CL. Articular mobility in an African population. Ann Rheum Dis. 1973;32(5):413-8.
29. Grahame R, Bird HA, Child A. The revised (Brighton 1998) criteria for the diagnosis of benign joint hypermobility syndrome (BJHS). J Rheumatol. 2000;27(7):1777-9.
30. Juul-Kristensen B, Rogind H, Jensen DV, Remvig L. Inter-examiner reproducibility of tests and criteria for generalized joint hypermobility and benign joint hypermobility syndrome. Rheumatology (Oxford). 2007;46(12):1835-41.

Appendix 2 – Exercise description

EXERCISE PROGRAMS FOR PARTICIPANTS IN THE COMPETE TRIAL

This document describes the details of the two exercise programs (Hip Exercise; HE, and Quadriceps Exercise; QE) that are compared in the COMPETE trial.

GENERAL CONSIDERATIONS

Both exercise programs are initiated at an individual clinical visit. An experienced physiotherapist introduces the trial participant to the exercise program that the participant has been allocated to (HE or QE) and provides instructions to the individual exercises.

Both exercise programs run for 12 weeks, with exercise sessions 3 times per week. Each training session is scheduled to last approximately 30 minutes. The exercise programs are home based with monthly supervision visits at the department of Physical and Occupational Therapy at Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark.

Focus of the exercises is on the quality of the performance – not quantity. It is considered very important to do the exercises with correct technique in order to gain as much as possible from the exercises. This is emphasised to the participants during the instruction session.
During the instruction session, the participants should be able to do at least one set of each exercise with satisfying quality (judged by the physiotherapist) before the participant is sent home to do the exercise program on him/her own.

Both programs consist of 3 resistance training exercises. The overall principles of the programs are based on the guidelines for strengthening exercises from the American College of Sports Medicine (1). The aim of the exercises are to gain muscular strength and through numerous neuromuscular mechanisms to enable greater force generation. Muscular strength is defined as the ability of a muscle or muscle group to exert a maximal external force.

The individual exercises in both exercise programs include concentric and eccentric muscle actions. Participants are instructed to perform the exercises in a moderate velocity, i.e. 1-2 seconds in the concentric movement and 1-2 seconds in the eccentric movement. Recommended load is 60-70% 1RM (repetition maximum (RM)) i.e. 8-12 repetitions. The recommended volume is 3 sets with recommended rest between sets of 1½ - 2 minutes (1). The exercise descriptors are summarized in Table 1.

Important note on RM estimation
Since we use elastic bands, free weights, and body weight as exercise resistance, it is not possible to estimate the exact repetition maximum. However, we inform the participants to approximate muscle fatigue within 8-12 repetitions in each set. An increase in resistance is recommended when the

participant is able to perform 2 repetitions more than the desired number (i.e. 14 or more) (1) with satisfying quality (progression principals are specified in the exercise description).

Exercise equipment
Elastic bands with different resistances are used for progression.
In addition it is recommendable to use a training mat or alternatively a large towel to lie on when necessary. For exercise QE-1 it is recommended to use a solid table or a high stool as a normal chair is typically too low for the legs to move freely.

Knee pain monitoring
The participants are instructed to consider their current knee pain on the day of exercise. A 0-10 numeric rating Scale (NRS, 0 equals no pain and 10 equals worst imaginable pain, figure 1) is used by the participant to assess their knee pain before, during and after each training session. If the participant experiences knee pain of an intensity of 6 or more before or during a session, the participant is instructed to decrease resistance and/or number of repetitions in that session. The pain ratings are recorded in the participants exercise diary at each session (appended).

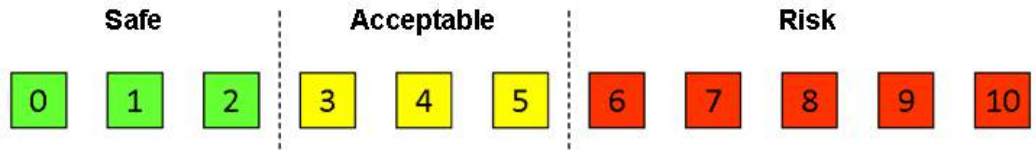


Figure 1: Visual numeric rating scale for pain monitoring ranging from 0 (no pain) to 10 (worst imaginable pain) with safe, acceptable, and risk pain zones indicated.

Both groups are instructed not to perform other activities that cause knee pain of 3 or more on the 0-10 NRS (figure above), except from when performing the exercises in which knee pain up to 5 is acceptable. If pain intensity exceeds 3 on the NRS, the participant is recommended to stop the activity. Participants are recommended to maintain daily activity level in the trial period to the extent that the pain is kept below this limit.

General instructions to participants in both groups
During activities of sports or daily life, participants are instructed to focus on good alignment. Good alignment is defined as keeping the anterior superior iliac spine (ASIS), the knee and the second toe in a straight line when moving, e.g. when ascending or descending stairs, bicycling, running, etc.

In the leaflet “Managing my patellofemoral pain” (Danish title: “Håndtering af mine forreste knæsmerter”, appendix J), information regarding alignment and malalignment is explicated in an easily read form.

THE HIP EXERCISE PROGRAM (HE)

The exercises used in the HE-group program (exercises HE-1, HE-2, and HE-3) have been chosen due to their documented activation of the hip abductors, external rotators, and hip extensors (2-5), are widely used in clinical practice, and because they do not strain the knee or patellofemoral joints that are painful in PFP patients. The exercises are easy to conduct and are easily progressed.

Warm up

Participants are instructed to warm-up by performing 20 repetitions of exercise HE-1 (see below) without external load.

HE-1 – SIDELYING CLAMP

Purpose: To progressively strengthen the hip abductors and external rotators.

Exercise description: Lay on the side with hips flexed approximately 30 degrees and knees flexed approximately 90 degrees. The pelvis should be kept stable and in neutral position throughout the exercise. Lift the upper knee as high as possible without pelvis tilt and any compensatory movements. Lower the knee again until the starting position. Load can be applied by placing an elastic band between the knees.

Focus: Main focus is on stabilizing the pelvis in a neutral position without tilting backwards when performing the exercise. In addition to make sure, that exertion is felt primarily in the gluteal muscles and not in the tensor fascia latea muscle, hamstrings or any other potential accessory muscle.

Progression: Progression is made by applying elastic bands with more resistance and/or applying more bands.

HE-2 SIDELYING HIP ABDUCTION

Purpose: To progressively strengthen the hip abductors

Exercise description: Lay on the side with the upper leg in full knee and hip extension (neutral position). The lower leg is kept with a slight hip flexion and above 90 degrees knee flexion during the exercise. The pelvis should be kept stable and in neutral position throughout the exercise. Lift the upper leg upwards and slightly backwards as high as possible without pelvis tilt and any compensatory movements. Lower the leg again until the starting position. Load can be applied by placing an elastic band between the legs. Progression is made by applying elastic bands with more resistance and/or applying more bands

Focus: Main focus is on stabilizing the pelvis in a neutral position without tilting backwards when performing the exercise. In addition to make sure, that exertion is felt primarily in the gluteus medius muscle and not in the tensor fascia latea muscle, hamstrings or any other potential accessory muscle.

Progression: After 4 weeks of training the exercise is progressed to standing as described in the following:

Exercise description: Stand on one leg with the pelvis and upper body in a neutral position and the knee on the stand-leg slightly flexed. The pelvis and upper body should be kept stable and in neutral position throughout the exercise. Lift the training-leg leg outwards and slightly backwards as far as possible without pelvis tilt and any compensatory movements. Return again to the starting position.

Focus: Main focus is on stabilizing the pelvis in a neutral position without rotating or tilting when performing the exercise. In addition to maintain good alignment between the feet, knees and anterior superior iliac spine in the weight bearing extremity (see patient information and guidelines).

Further Progression: Load can be applied by placing an elastic band between the legs. Progression is made by applying elastic bands with more resistance and/or applying more bands

HE-3 PRONE/STANDING HIP EXTENSION

Purpose: To progressively strengthen the hip extensors

Exercise description: Lie on your stomach on a table top, positioned with your legs off the end of the table and hips flexed to approximately 90 degrees with the knees flexed and the feet placed on the floor. Extend one leg at a time, maintaining flexion of the knee. Do not allow the leg to abduct during hip extension. Return again to the starting position.
Emphasis is placed on extension at the hip, avoiding extension of the spine. The lower back should be kept in neutral position with a slight lordosis and the pelvis should be kept stable without tilting or rotating throughout the exercise.
Alternative exercise set-up: If it is not possible to find a suitable table to lie on, you can do the exercise standing in a bend over position equal to the above description. Instead of lying on a table top, you can support the upper body by holding on to a chair or table.

Focus: Main focus is on maintaining a natural posture with a slight lordosis in the lower back and to avoid compensatory movements in the spine or pelvis.

Progression: Load can be applied by placing an elastic band between the legs from under the standing foot to the knee of the moving limb. Progression is made by applying elastic bands with more resistance and/or applying more bands.

THE QUADRICEPS EXERCISE PROGRAM (QE)

The exercises used in the QE-group program (exercises QE-1, QE-2, and QE-.3) are also widely used in clinical practice. The exercises has been shown effective in recruiting the quadriceps muscle (PMID: 18560185, 22310511) and are effective in treating PFP (PMID: 24766358, PMID: 17469667) . Since two of the exercises (QE-2 and QE-3) are multi-joint exercises, several muscles are recruited when performing the movement. In the weight bearing squat (QE-2) and lunge (QE-.3), the quadriceps muscles are activated in concomitance with primarily the hamstrings and the gluteal muscles. Nevertheless we chose these exercise because they are widely used and thus fit well into the pragmatic nature of the COMPETE trial.

Warm up
Participants are instructed to warm-up by performing 20 repetitions of exercise QE-1 (see below) without external load.

QE-1 SITTING KNEE EXTENSION

Purpose: To progressively strengthen the knee extensors (open kinetic chain).

Exercise description: Sit on a table or similar with the knees bended and the lower extremity hanging freely. The upper body must be in an upright or slightly reclined position. Extend the knee from 90 degrees of knee flexion until full knee extension. Lower the leg again until start position. Keep the foot in a dorsiflexed position throughout the exercise.
Load can be applied by placing an elastic band between the feet and fixing the untrained leg in knee flexion.

Focus: Main focus is on stabilizing the pelvis in a neutral position without retroverting when performing the exercise. In addition to make sure, that the exercise is performed in the desired range of motion, i.e. from 90 degrees of knee flexion until full knee extension.

Progression is made by applying elastic bands with more resistance and/or applying more bands.

QE-2 SQUAT

Purpose: To progressively strengthen the knee extensors, hamstring muscles and gluteal muscles involved in the squatting movement.

Exercise description: Stand with a shoulder width distance between the feet. Flex the knees while maintaining good posture in the upper body and good alignment in the lower extremities until 90 degrees of knee flexion. Straighten the knees again until the starting position.
Load can be applied by adding weight in a backpack (e.g. sand, flour, bottles of water) or by holding dumbbells in the hands.

Focus: Main focus is on maintaining good alignment between the feet, knees and anterior superior iliac spine (see patient information and guidelines). In addition to make sure, that the lower back is kept stable in a slight lordosis throughout the movement.

Progression is made by adding weight. Regression is made by flexing the knees less than 90 degrees, e.g. 45 degrees or 70 degrees.

QE-3 FORWARD LUNGE

Purpose: To progressively strengthen primarily the knee extensors, hamstring muscles and gluteal muscles.

Exercise description: Stand with one foot in front of the other with the distance of a large step between the feet. The heel of the back foot is kept slightly raised throughout the exercise. Bend the knees so that the back knee touches or almost touches the floor and the front knee is in a 90 degrees flexion. Extend the knees again until the starting position.

Focus: Main focus is on maintaining good alignment between the front foot, knee and anterior superior iliac spine (see patient information and guidelines). The front knee is not allowed to exceed the vertical line from the first toe. In addition to make sure, that the lower back is kept stable in a slight lordosis and the upper body is kept in an upright position throughout the movement.

Progression: Load can be applied by adding weight in a backpack (e.g. sand, flour, bottles of water) or by holding dumbbells in the hands. Progression is made by adding weight. Regression is made by flexing the front knee less than 90 degrees, e.g. 45 degrees or 70 degrees.

Table 1. Exercise descriptors

	HIP EXERCISES				QUADRICEPS EXERCISES		
	HE-1	**HE-2**	**HE-2 progression**	**HE-3**	**QE-1**	**QE-2**	**QE-3**
1. Load magnitude	8-12 RM.	8-12 RM.	8-12 RM.	8-12 RM.	8-12 RM.	8-12 RM.	8-12 RM.
2. Number of repetitions	8-12 reps.	8-12 reps.	8-12 reps.	8-12 reps.	8-12 reps.	8-12 reps.	8-12 reps.
3. Number of sets	3 sets	3 sets	3 sets	3 sets	3 sets	3 sets	3 sets
4. Rest in between sets	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.	1 min. 30 sec. – 2 min.
5. Number of exercise interventions per week	3/week	3/week	3/week	3/week	3/week	3/week	3/week
6. Duration of the experimental period	12 weeks	12 weeks	12 weeks	12 weeks	12 weeks	12 weeks	12 weeks
7. Fractional and temporal distribution of the contraction modes per repetition and duration of one repetition	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric	1-2 s concentric, 1-2 s eccentric
8. Rest in-between repetitions	No	No	No	No	No	No	No
9. TUT	2-4 s/repetition	2-4 s/repetition	2-4 s/repetition	2-4 s/repetition	2-4 s/repetition	2-4 s/repetition	2-4 s/repetition
10. Volitional muscle failure	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11. Range of motion	See exercise description	See exercise description	See exercise description	See exercise description	90 degrees to 180 degrees	See exercise description	See exercise description
12. Recovery time in-between exercise sessions	48 h	48 h	48 h	48 h	48 h	48 h	48 h
13. Anatomical definition of the exercise	See exercise description	See exercise description	See exercise description	See exercise description	See exercise description	See exercise description	See exercise description

1. American College of Sports M. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.
2. Boren K, Conrey C, Le Coguic J, Paprocki L, Voight M, Robinson TK. Electromyographic analysis of gluteus medius and gluteus maximus during rehabilitation exercises. Int J Sports Phys Ther. 2011;6(3):206-23.
3. Distefano LJ, Blackburn JT, Marshall SW, Padua DA. Gluteal muscle activation during common therapeutic exercises. J Orthop Sports Phys Ther. 2009;39(7):532-40.
4. Ekstrom RA, Donatelli RA, Carp KC. Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises. J Orthop Sports Phys Ther. 2007;37(12):754-62.
5. MacAskill MJ, Durant TJ, Wallace DA. Gluteal muscle activity during weightbearing and non-weightbearing exercise. Int J Sports Phys Ther. 2014;9(7):907-14.

Appendix 3 – Information leaflet (Danish)

VIGTIGE FAKTORER OG BEHANDLINGSMULIGHEDER

(Se nærmere detaljer indeni)

VIGTIGE FAKTORER SOM MAN SKAL HAVE FOKUS PÅ

1. Nedsat funktion af og svag hoftemuskulatur.
2. Nedsat funktion af og svag lårmuskulatur.
3. At man falder indad på foden (øget pronation)

VIGTIGE BEHANDLINGSMULIGHEDER

1. Øvelser der øger styrke og funktion af hofte- og lårmuskulaturen.
2. Tapening af knæskallen for at reducere smerten på kort sigt.
3. Indlægssåler med svangstøtte hvis du falder meget indad på foden (øget pronation)

TILTAG FOR AT HÅNDTERE DINE FORRESTE KNÆSMERTER

1. Hvis du tror, at du har forreste knæsmerter bør du hurtigst muligt søge hjælp – dette vil øge dine chancer for en succesfuld genoptræning
2. Der er mange effektive behandlingsmuligheder som du kan drøfte med din behandler.
3. Effekten af din genoptræning vil øges hvis du yder en aktiv indsats.
4. Justering af dit fysiske aktivitetsniveau er ofte det første skridt mod en succesfuld genoptræning.
5. Sørg for at du langsomt og sikkert øger dit fysiske aktivitetsniveau (figur 5)

Figur 5: Gradvis og sikker opbygning af aktivitetsniveau og sportsdeltagelse

Dette informationsmateriale er fremstillet af Dr. Christian Barton og Dr. Michael Rathleff, som en del af et ikke-kommercielt projekt.

HÅNDTERING AF MINE FORRESTE KNÆSMERTER

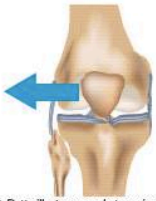
Patellofemorale smerter, også kaldet forreste knæsmerter (smerter omkring eller bag knæskallen) er meget hyppigt og ses hos mænd og kvinder på alle aktivitetsniveauer. Patellofemorale smerter resulterer i knæsmerter under almindelige dagligdagsaktiviteter såsom almindelig gang, løb, når man sidder, samt når man går på trapper.

Der er mange årsager til patellofemorale smerter, og derfor mange forskellige behandlingsmuligheder. Informationen i denne folder kan hjælpe dig med den mest hensigtsmæssige behandling for dine knæsmerter. Det anbefales at du tager kontakt til en sundhedsprofessionel for yderligere information om behandling af dine knæsmerter.

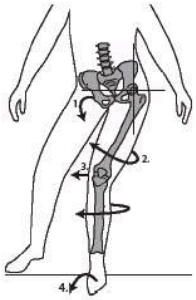
HVAD KAN VÆRE ÅRSAGERNE TIL MINE KNÆSMERTER?

Øget belastning eller varieret og hurtig stigning i fysisk aktivitetsniveau som dit knæ ikke kan klare (figur 1) er sandsynlig medvirkende årsag til udvikling af smerter. Uhensigtsmæssige bevægemønstre kan også medvirke til at knæskallen bevæger sig mod ydersiden af knæet (Figur 2), og dermed ikke bevæger sig normalt i knæets føringsfure. Flere faktorer kan føre til denne uhensigtsmæssige bevægelse af knæskallen (figur 3). Der er også andre faktorer som kan bidrage til udviklingen af forreste knæsmerter f.eks. knæleddets opbygning, tidligere skader, tidligere operationer og andre sygdomstilstande. Disse faktorer kan du drøfte med din behandler.

Figur 1: Hurtig øgning i aktivitetsniveau og sportsdeltagelse kan føre til forreste knæsmerter



Figur 2: Dette illustrerer ændret sporing af knæskallen



Figur 3: Bevægelser som kan føre til ændret sporing af knæskallen

1. Bækkenet tilter nedad på modsatte side hvilket øger spændingen på ydersiden af benet. Dette medfører at knæskallen trækkes mod ydersiden.
2. Hoften falder sammen pga. svaghed samt dårlig funktion af hoftemuskulaturen. Herved indadroterer lårbenet under knæskallen.
3. Lårets muskulatur er svag eller fungerer utilstrækkeligt og fører til manglende støtte af knæ og knæskal.
4. Foden falder indad hvilket medfører at skinneben og knæ roterer indad under knæskallen.

BEHANDLINGSMULIGHEDER

(Kvalitetsøvelser er nøglen til god genoptræning)

TRÆNING

HVILKE BEHANDLINGSPRINCIPPER ER VIGTIGE?

1. Din behandler foreslår muligvis en periode med ro før opstart af øvelser.
2. I starten kan siddende eller liggende øvelser hjælpe med til at hofte og lårmuskler kan trænes uden smerter.
3. Når smerten tillader det, bør øvelserne udføres i stående stilling der efterligner hverdagsaktiviteter.
4. Din behandler bør supervisere dig i nye øvelser for at sikre korrekt teknik.
5. Brugen af spejle og videooptagelser kan hjælpe dig til korrekt teknik når du udfører øvelserne derhjemme.
6. Øvelserne vil ikke hjælpe dig medmindre teknikken er korrekt og at de udføres regelmæssigt.

HVILKE ØVELSER SKAL JEG UDFØRE?

1. Svaghed og dårlig funktion af hofte og lårmuskler er meget almindelig, så du skal formentlig lave øvelser for at forbedre dette.
2. Af og til kræves der også øvelser til foden eller ryggen.
3. Du skal muligvis udspænde læg, hase eller lårets forside.
4. Øvelsernes sværhedsgrad skal øges hen imod aktiviteter der tidligere gav smerter (gå i hug, trappegang, løb osv.). Dette er for at sikre gode bevægelsesmønstre under udførelsen (se eksempel i figur 4).

ANDRE BEHANDLINGER

SMERTEREDUKTION

1. Brug af tape kan reducere smerter på kort sigt – din behandler kan lægge det på eller lære dig hvordan det gøres.
2. En bandage kan også mindske smerter.
3. Indlægssåler kan nogle gange reducere dine knæsmerter - din behandler kan vurdere hvilke indlæg der er hensigtsmæssige for dig eller henvise dig til én som kan.

4a. Dårlig kontrol 4b. Forbedret kontrol af af hofte og bækken hofte og bækken

Figur 4 Gang ned ad trapper

HVORDAN KAN MIN BEHANDLER ELLERS HJÆLPE?

1. Vejlede dig i de mest hensigtsmæssige øvelser samt andre behandlinger.
2. Besvare øvrige spørgsmål vedrørende dine knæsmerter og ved behov forklare dig i detaljer de informationer, der er beskrevet i denne folder.
3. Hjælpe dig til at forstå hvorfor du har knæsmerter, hvilke faktorer der med størst sandsynlighed har givet dig smerter og hvordan du kan ændre dine aktiviteter for at reducere dine knæsmerter og sikre god genoptræning.
4. Tilbyde manuel behandling som muligvis kan være vigtig for at forbedre dine smerter og bevægelighed.

Appendix 4 – Statistical Analysis Plan (SAP) Study 1

Bispebjerg og Frederiksberg Hospital

STATISTICAL ANALYSIS PLAN

Comparative efficacy of therapeutic hip and knee exercise for patellofemoral pain: a pragmatic randomised trial.

Trial Registration

Health Research Ethics Committee Number: H-16045755 (approved December 15, 2016)

Clinicaltrials.gov Trial registration identifier: NCT03843931

Internal protocol number: FYS-2016-004

Protocol Version and Date

This document has been written based on information contained in the study protocol version 1.2

April 18, 2017

Statistical Analysis Plan Version and Date

Version 1

October 28, 2021

Statistical Analysis Plan Authors

Marius Henriksen

Rudi Neergaard Hansen

CHANGE HISTORY

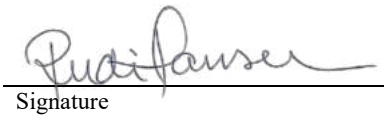
Protocol version	Updated SAP version	Section Number Changed	Description of and reason for change	Date changed

1 SIGNATURES

Approved by Principal Investigator Marius Henriksen

Marius Henriksen, Prof, PhD The Parker Institute	 Signature	29-10-2021 Date

Approved by Trial Manager Rudi Neergaard Hansen

Rudi Neergaard Hansen, Msc. Bispebjerg and Frederiksberg Hospital	 Signature	29-10-2021 Date

2 PURPOSE

This statistical analysis plan (SAP) describes detailed aspects of data preparation and analysis and was set up before starting the final analysis. The SAP is based on the final trial protocol (Version 1.2, April 18, 2017).

3 STUDY SYNOPSIS

Background and rationale:	Patellofemoral Pain (PFP) is a common knee problem, which particularly affects adolescents and young adults. PFP is characterised by significant retropatellar and/or peripatellar pain and impairment of function and quality of daily life. Exercise programs targeting either hip or knee muscles are recommended, but it is unclear if these exercise programs produce equivalent results. The purpose of this study is to compare changes in pain and function for patients assigned to a focused "Quadriceps Exercise" protocol or a "Hip Exercise" protocol for a 12-week period.
Objectives:	*Primary objective:* To assess efficacy equivalence between a focused "Quadriceps Exercise" protocol (QE) and a "Hip Exercise" protocol (HE), on changes in knee pain and function in individuals with PFP in the short term (12 weeks). *Secondary objectives:* 1) To assess efficacy equivalence between QE and HE on changes in knee pain and function in individuals with PFP in the long term (26 weeks), 2) to assess efficacy equivalence between QE and HE on: The 5 subscales of the KOOS questionnaire (Pain, Symptoms, Function, Sports/Recreation, Quality of Life), isometric muscle strength of hip abductors, hip external rotators, hip extensors, and quadriceps, Dynamic Assessment of Pain, Pain Self-Efficacy Questionnaire, EuroQoL EQ-5D-3L Questionnaire, and Transition Questionnaire of global perceived effect on overall health, pain, and function.
Outcomes:	*Primary outcome:* Change from baseline in the KUJALA scoring questionnaire at week 12. *Key secondary outcomes:* - Change from baseline in the KOOS pain subscore at week 12 - Change from baseline in the KOOS function subscore at week 12 - Change from baseline in the KOOS quality of life subscore at week 12 *Other secondary outcomes:* Change from baseline in the KOOS Symptoms and Sports/Recreation subscores at week 12, change from baseline in isometric muscle strength of hip abductors, hip adductors, hip external rotators, hip internal rotators, hip extensors, hip flexors, knee flexors (hamstrings), and knee extensors (quadriceps) at week 12, change from baseline in Dynamic Assessment of Pain at week 12, change from baseline in Pain Self-Efficacy Questionnaire at week 12, change from baseline in EuroQoL EQ-5D-3L Questionnaire at week 12, The Transition Questionnaire of global perceived effect on overall health, pain, and function at week 12, change from baseline in the outcomes measured at week 26 (only questionnaire data).
Study design:	The trial is a randomised, controlled, equivalence trial with two parallel groups comparing QE and HE.
Statistical considerations:	Primary analyses will be based on an intention-to-treat (ITT) principle. Continuous scores will be analysed using mixed linear models adjusted for baseline values of the scores. As secondary analysis we will repeat the primary analysis on the per protocol (PP) population. If the ITT and PP analyses agree, confidence in a potential equivalence claim is increased. Sensitivity analyses will be done using a data set with missing data replaced using multiple imputation, and adjusting for intervention adherence. Adverse events will be presented in a descriptive way for both groups.

For further details regarding the trial design, please see the protocol version 1.2, April 18, 2017.

4 INTRODUCTION

4.1 Background and rationale

Patellofemoral pain (PFP) is a common knee problem, which particularly affects adolescents and young adults. PFP is characterised by significant retropatellar and/or peripatellar pain and impairment of function and quality of daily life. Exercise has repeatedly been shown beneficial for pain and physical function and is unequivocally recommended as a core component of the management of PFP. Different types of exercise (e.g., quadriceps strengthening, hip strengthening and functional/neuromuscular exercises) have been investigated. In general, these different types of exercises produce similar small to moderate beneficial effects in pain and physical function. Evidence has been accumulating to support the importance of quadriceps and hip muscle control and strengthening in the treatment of PFP, but studies including direct comparisons of the separate treatment protocols are few (1-5), and intervention durations and follow-up periods have been short (i.e., 6-12 weeks). Furthermore, even though different strengthening regimens have been compared, claims of equivalence cannot be established from available studies as nonsignificant superiority tests only in very rare occasions can be interpreted as proof of no difference between the two treatments. Tests of equivalence normally require an established gold standard treatment against which a new treatment is tested (for equivalence). However, in exercise for PFP, neither hip nor quadriceps exercise programs are considered gold standard, wherefore we aim to assess if the two types of exercise are non-inferior to each other. This is done through a randomised trial designed to test for equivalence of the two exercise programs.

Accordingly, the purpose of this study is to assess efficacy equivalence between a focused "Quadriceps Exercise" protocol (QE) and a focused "Hip Exercise" protocol (HE) in pain and function in patients with PFP.

4.2 Study Objectives

The primary objective of this trial is to assess efficacy equivalence between QE and HE on changes in knee pain and function in individuals with PFP.

The secondary objectives are to compare the QE and HE on the following

- Changes in patient-reported physical function, knee symptoms, quality of life, and participation in sports and recreation

- Change in physical performance
- Changes in patients perceived overall effect

5 STUDY METHODS

5.1 Trial Design

The trial is a single centre, randomised, parallel-group, 26 weeks (6 months), equivalence trial comparing a 12-weeks focused "Quadriceps Exercise" protocol and a 12-weeks focused "Hip Exercise" protocol with a primary endpoint at 12 weeks (after treatment) and a follow-up at 26 weeks.

The trial is conducted among patients with PFP. A total of 200 patients has been randomly assigned on a 1:1 basis to one of the two treatments, QE or HE.

5.2 Randomization

Eligible participants have been randomly assigned - in permuted blocks of 4 and 6 - according to a computer-generated list of random numbers, to one of the two groups (QE or HE).

The randomisation is equal (1:1), meaning that 100 participants are allocated to each group. A coded randomisation list was available to the clinical staff administering the interventions.

5.3 Blinding

Investigators, study coordinators, clinical staff, study staff, and other personnel directly involved in the study, are blinded to the group allocation until all primary and secondary analyses are completed. Participants and staff involved in the exercise delivery are not blinded to the group allocation. Information that could potentially unblind otherwise blinded staff will not be shared and will be stored in facilities with limited access until the study is completed. Unblinding of blinded personnel does not preclude the related participants' continued participation in the study.

5.4 Sample Size and Power

The sample size has been calculated to test the equivalence of the QE and HE programs in the assessment of change in the KUJALA questionnaire. With 77 participants per group, the study will have 90% power assuming the expected group difference in mean changes from baseline is 0, the common standard deviation is 15 (0-100 scale), with a delta (equivalence margin) of 8 units (0-100 scale) corresponding to the suggested minimum clinically relevant difference, and a significance level of 5%. With an expected drop-out during the study we will randomise and allocate 200 participants (100 to each group); analysing the PP population.

5.4.1 Statistical power calculation for potential superiority claim

A sample size of 200 in total will provide strong statistical power to detect group differences in favour of either of the two investigational treatments.

For a two-sample pooled t test of a normal mean difference with a two-sided significance level of 0.05 (P<0.05), assuming a common standard deviation of 15 KUJALA points, a total sample size of 200 assuming a balanced design has a power of 80.4% to detect a mean difference of 6 KUJALA-Points (corresponding to a small effect size of 0.4).

5.5 Framework

This is an equivalence trial.

5.6 Statistical Interim Analyses and Stopping Guidance

No statistical interim analysis has been planned and there is no guidance for stopping the trial.

5.7 Timing of Final Analysis

Final analysis will take place in one stage: The first (and main) report/publication of the trial will be prepared for the QE/HE comparison when every trial participant has reached 26 weeks follow-up and data for the primary and secondary outcomes have been received and cleaned (anticipated to be March 2022).

5.8 Timing of Outcome Assessments

The schedule of study procedures and visit windows are given in the Table 1. The start time (Day 1) is the scheduled day of the participant's first treatment.

			Week												
	Screening	Baseline	1	2	3	4	5	6	7	8	9	10	11	12	26
Day	-56 to -28	-28 to 0	1-7	8-14	15-21	22-28	29-35	36-42	43-49	50-56	57-63	64-70	70-77	78-84	182
Written information	●														
Oral information	●														
Procedure															
Eligibility criteria	●														
Informed consent		●													
Randomisation		●													
Interventions															
QE			●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	
HE			●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●	
Outcomes															
Questionnaires															
Kujala		●												●	●
KOOS		●												●	●
Pain self-efficacy questionnaire		●												●	●
EuroQoL EQ-5D-3L Questionnaire		●												●	●
Transition Questionnaire of global perceived effect		●												●	●
Clinical assessment															
Isometric strength		●												●	
Dynamic assessment of pain		●												●	

6 OUTCOMES

6.1 Study knee

At inclusion a study knee was selected, which is subject to all subsequent assessment:

- The study knee will be defined as the symptomatic knee with a diagnosis of PFP
- If both knees are eligible, the more symptomatic knee will be selected (selected by participant)
- If both knees shave equivalent symptoms (reported by participant), the study knee will be randomly assigned.

6.2 Primary outcome

The primary outcome is assessed at week 12 as change from baseline in the KUJALA questionnaire – a widely used and well-validated survey instrument evaluating pain and function in PFP patients. We will analyse the group difference in the mean changes from baseline in the KUJALA questionnaire in the study knee between QE vs HE after 12 weeks.

6.3 Key Secondary outcomes

The following outcome is assessed as key secondary outcome:

- Change from baseline in the KOOS pain subscore at week 12
- Change from baseline in the KOOS function subscore at week 12
- Change from baseline in the KOOS quality of life subscore at week 12

6.4 Other secondary outcomes

The following outcomes are assessed as other secondary outcomes:

- Change from baseline in the KOOS Symptoms and Sports/Recreation subscores at week 12

- Change from baseline in isometric muscle strength of hip abductors, hip adductors, hip external rotators, hip internal rotators, hip extensors, hip flexors, knee flexors (hamstrings), and knee extensors (quadriceps) at week 12
- Change from baseline in Dynamic Assessment of Pain at week 12
- Change from baseline in Pain Self-Efficacy Questionnaire at week 12
- Change from baseline in EuroQoL EQ-5D-3L Questionnaire at week 12
- The Transition Questionnaire of global perceived effect on overall health, pain, and function.
- Change from baseline in the outcomes measured at week 26 (only questionnaire data)

6.5 Definition of outcome variables

6.5.1 The KUJALA score questionnaire

The KUJALA score questionnaire - sometimes called the Anterior Knee Pain Scale - is a disease specific validated disability scale ranging from 0 (complete disability) to 100 (fully functional). It is a 13-item self-report questionnaire that documents response to 6 activities (walking, running, jumping, climbing stairs, squatting, and sitting for prolonged periods with knees bent), as well as symptoms such as limp, inability to weight bear, swelling, abnormal patellar movement, muscle atrophy, and limitations in knee flexion. The minimal clinical important difference is reported to range from 8 to 19 points (6) in patients with PFP.

6.5.2 KOOS

The Knee injury and Osteoarthritis Outcome Score (KOOS), a disease-specific instrument designed to assess health related quality of life (QoL) in patients with knee injuries. The KOOS consists of 42 items covering five domains, namely, *Pain* (9 items), *Function* (in Activities of Daily Living) (17 items), *Knee-related QoL* (4 items), *Symptoms* (7 items), and *Sports and Recreation* (5 items). The KOOS uses a five-point Likert scale scoring system (ranging from 0 (least severe) to 4 (most

severe)). The minimal clinical important difference in the subscores is suggested to be 8-10 points in patients with knee osteoarthritis (www.koos.nu[1]).

We will calculate all KOOS domains from the questionnaire values as outlined in the user guide (www.koos.nu[1]). Normalized scores are calculated for each domain with 100 indicating no symptoms and functional impairment and 0 indicating extreme symptoms and functional impairment. If the number of missing items is less than or equal to 2 in a subscale, they will be substituted by the average item value for that subscale. If more than two items of the subscale are omitted the response will be considered invalid and no subscale score calculated.

6.5.3 Isometric muscle strength

Isometric muscle strength of hip abductors and adductors, hip external and internal rotators, hip flexors and extensors, quadriceps, and hamstrings is performed by using a handheld dynamometer and is measured in Newtons (N). Measurement variation has previously been assessed and found to be less than 5% when assessing hip abduction, hip external rotation and knee extension, and less than 10% when assessing knee flexion (7). The muscle strength tests are conducted following published testing protocols (7-9). Among healthy individuals, the minimal detectable changes are reported to be: 45.9 N for hip extensors, 27.1 N for hip flexors, 21.6 for hip adductors, 20.5 N for hip internal rotators, 11.6 N for hip adductors, 9.7 N for hip external rotators (7), 24.6 N for knee flexors and 18.2 N for knee extensors (9). Since no minimal clinical important differences exist for muscle strength, we set equivalence margins to one half standard deviation of the pooled baseline values in this trial. Hence the exact MCID values cannot be listed a priori as the database has not been locked when this statistical analysis plan was written.

6.5.4 Dynamic Assessment of Pain

The Functional Weight Bearing Pain Test is a simple performance test with an integrated pain score, designed to provide useful information for monitoring treatment progress and evaluating treatment effects in clinical physiotherapy practice. The patient is asked to perform as many squatting movements (both legs) as possible within 30 seconds. The knees should reach approximately 90 degrees of flexion and full extension for each squat. This is supervised by the

[1] Accessed on October 23rd, 2021

investigator. The outcome of the test is the knee pain during the test on a 0-10 Verbal Rating Scale (VRS) rated immediately after the test. The minimal clinical important difference is suggested to be 2.4 points in patients with knee osteoarthritis (10).

The test takes about 1 minute to perform including instructions and does not require any equipment besides a stopwatch/watch. The result is a direct measure of the patient's ability to perform a repeated movement within a short timeframe and for the degree of pain during a weight bearing movement, which reflects the limitations of daily activities due to PFP.

6.5.5 Pain Self-Efficacy Questionnaire

The pain self-efficacy questionnaire is a 10-item questionnaire developed to assess the confidence people with pain have in performing activities while in pain. It is applicable to all persisting pain presentations and covers a range of functions. Confidence in performing activities is rated on a 7-point (0-6) Likert scale with 0 representing not at all confident and 6 representing completely confident. A total score is calculated by summing the answers producing a score between 0 and 60. Higher scores reflect stronger self-efficacy beliefs. MCID is reported to be 5.5 in patients with chronic low back pain (11), however, in a Danish population of low back pain patients, the smallest detectable change (SDC) was reported to be 12.67 points (12). Neither MCID nor SDC is reported for patients with knee pain. We define the MCID as one half standard deviation of the pooled baseline values in this trial. Hence the exact MCID used for the analyses cannot be listed a priori as the database has not been locked when this statistical analysis plan was written.

6.5.6 EuroQoL EQ5D Questionnaire

EQ5D is a standardised patient-reported instrument for use as a measure of health outcome and quality of life. The answers to the five domain statements can be translated into a single index value within the range of 1,000 to -0.624 using so-called preference weights based on Danish normative data, as higher values indicate better health-related quality of life and vice versa. The MCID is reported to be 0.32 points in patients with hip or knee osteoarthritis (13).

6.5.7 Transition Questionnaire of global perceived change in overall health, pain, and function

In transition ratings the participants are asked at follow-up to compare their current state with the state at baseline. This approach has limitations regarding recall bias and influence of numerous known and unknown parameters. However, a combination of changes on current state ratings (KUJALA or KOOS) and a transition questionnaire (TRANSQ) may enhance the interpretation of the results of the study.

We have designed a transition questionnaire on which the participants initially answer if their current state is “unchanged, worse” or “better” compared to the baseline visit. An “unchanged” equals a transition score of 0. If the participant answers “worse”, he/she is asked to rate the degree of worsening on a 7-point Likert scale, and the corresponding scores range from -1 to -7. Correspondingly, if a participant answers “better”, he/she is asked to rate the degree of improvement on a 7-point Likert scale, and the corresponding scores range from 1 to 7. Thus, the Transition score range from -7 (worsening) to 7 (improvement), with the mid-point – 0 – representing no change. The transition scale is used to assess overall knee related health status. A score of 3 points is considered clinically meaningful (14).

6.6 Adverse and serious adverse events

The investigators and clinical staff monitor each participant for evidence of adverse events (AEs) throughout the study. The investigator will assess and record any AE in detail including the date of onset, description, severity, duration and outcome, relationship to study treatment, and any action(s) taken.

An investigator will adjudicate all reported AEs based on available and relevant medical records.

7 DATA MANAGEMENT

7.1 Data validation

All variables used in the database, including derived variables, will be checked for missing values, outliers and inconsistencies. We do not expect many faulty data points because error checks and warnings were implemented into the eCRF (Redcap).

7.2 Data preparation

7.2.1 Changes from baseline

The primary outcome is change from baseline in KUJALA questionnaire at week 12. This will be calculated for each individual as the baseline value subtracted from the week 12 value:

$$KUJALA_{change\_week12} = KUJALA_{week12} - KUJALA_{baseline}$$

Thus, a positive change value indicate that the week 12 value is greater than the baseline value, which suggest an improvement in the KUJALA score (= less pain and higher function).

The same calculation will be applied for the outcomes defined as change from baseline at various time points in the trial:

$$VARIABLE_{change\_week\textit{i}} = VARIABLE_{week\_\textit{i}} - VARIABLE_{baseline}$$

The interpretation of calculated change values are as follows:

OUTCOME	INTERPRETATION OF POSITIVE CHANGE VALUE
KUJALA (primary outcome)	Improvement
KOOS all subscales	Improvement
Isometric muscle test	Improvement
Dynamic Assessment of Pain (repetitions)	Improvement
Dynamic Assessment of Pain (pain)	Worsening
Pain Self-Efficacy Questionnaire	Improvement
EuroQoL EQ5D Questionnaire	Improvement

8 TRIAL POPULATIONS

8.1 Participant flow

A CONSORT participant flow diagram will be drawn following the CONSORT standards (see Shell Figure 1).

The flow diagram will be used to summarise the number of patients who were:

- assessed for eligibility at screening
- ineligible at screening*
- eligible but not randomised*
- received the randomised allocation

- did not receive the randomised allocation*
- lost to follow-up at week 12 and 26*
- withdrawals at week 12 and 26*
- discontinued the intervention*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

*reasons will be provided.

8.2 Intention-To-Treat population

The Intention-To-Treat (ITT) population consist of all randomized patients irrespective of whether the patient actually received study intervention or the patient's compliance with the study protocol, in the treatment group to which the participant was assigned at randomisation (Intention-To-Treat principle). A patient will be considered randomised as soon as a treatment is assigned according to the allocation sequence.

8.3 Per Protocol Population

The per protocol (PP) population consists of all participants in the ITT population who did not have any major protocol deviations that could make the interpretation of analyses on the ITT population difficult.

The following are pre-defined major protocol violations with a direct bearing on the primary outcome:

- Not adherent to the allocated intervention (see below for definition of satisfactory adherence)
- Initiation of other exercise programs/regimes than the one the participants are allocated to during the main trial phase (week 1-12).
- Surgery to the lower extremity during trial participation
- Failure to perform primary endpoint assessment, i.e. KUJALA questionnaire not assessed at week 12
- Early discontinuation of trial participation (before week 3)
- Week 12 visit not completed within +/- 7 days of the specified time window
- Non-compliance with any of the eligibility criteria

The number (and percentage) of patients with major protocol deviations will be summarised by treatment group with details of type of deviation provided. The number of randomised participants in each group will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

Non-pharmacological treatments and habitual use of pharmacological therapies are allowed. The usage of such treatments/therapies will be recorded on the case report form (CRF) and reported in a descriptive way for both groups.

8.4 Satisfactory adherence

Adherence to the prescribed exercise protocol is monitored by a self-administered exercise diary. The participants are asked to record date, number of repetitions and sets for each exercise, and the resistance (i.e., elastic band colour corresponding to a specified resistance or weights in kg.) for each exercise session.

Adherence is assessed based on the percent of the scheduled number of training sessions that was performed. A training session is considered performed, if an exercise activity is registered at a given date, even if the repetitions, sets or exercises are only partly recorded. The number of scheduled training sessions for both intervention groups is predefined in the trial protocol and equals 36 sessions for 12 weeks.

The following pre-defined criteria for satisfactory intervention adherence have been set: Have performed at least 24/36 of the scheduled training sessions (66%).

Descriptive statistics on the percent compliance (Mean, SD) will be summarized by randomisation group. Also, the number and % of participants receiving at least 66% of the prescribed treatment will be presented by treatment group.

8.5 Safety population

The safety population consists of all participants in the ITT population who has completed at least 1 exercise session.

9 STATISTICAL ANALYSES

9.1 General

In the primary analysis, all participants will be analysed using the ITT population according to the intention-to-treat principle.

Neither ITT nor PP analyses have perfect properties in equivalence studies. Therefore, current recommendations state that both ITT and PP analysis should be done and support each other for equivalence to be claimed. The underlying principle is that when ITT and PP provide identical conclusions, the confidence level of the investigator for the study results is augmented. We choose the ITT as primary analysis because it preserves the advantages of randomisation and is less prone to selection bias than PP. Further, the validity of PP analyses depends on assumptions about confounding that cannot be empirically verified. Further, we employed two treatments that require several treatment sessions (opposed to baseline all-or-nothing interventions or single-intervention studies). True PP analyses would require perfect protocol adherence (100%), which is unrealistic and therefore, we a priori defined a "satisfactory adherence" defining the PP population (see section 8.4). Such threshold is arbitrary and therefore debatable and hence we believe the ITT population to be the better choice as primary analyses. The PP analyses will be secondary and used to assess the robustness of the primary ITT analyses.

We will not apply explicit adjustments for multiplicity, rather we will analyse the key secondary outcomes in a prioritized order (i.e. "inverse gatekeeping procedure"): The analyses of the secondary outcomes will be performed in sequence until one of the analyses fails to show equivalence.

The hierarchy of the secondary outcomes including equivalence margins are as follows:

1. Change from baseline in the KOOS pain subscore at week 12
2. Change from baseline in the KOOS function subscore at week 12
3. Change from baseline in the KOOS quality of life subscore at week 12

All other secondary outcomes will be analysed, i.e. no hierarchy applied.

The statistician will be blinded to the treatment allocation at the time of the primary analysis of primary and secondary outcomes. Once the primary analysis is accomplished, the statistician may be unblinded.

9.2 Equivalence margins

As this is an equivalence trial, the following equivalence margins has been set prior to the analyses. *Equivalence* will be claimed if the computed 95% confidence interval of the estimated group difference in an outcome does not include the below equivalence margins.

OUTCOME MEASURE	EQUIVALENCE MARGINS
Primary outcome	
KUJALA questionnaire at 12 weeks	± 8 points
Key Secondary outcomes	
KOOS pain subscore at week 12	± 8 points[1]
KOOS function subscore at week 12	± 8 points[1]
KOOS quality of life subscore at week 12	± 8 points[1]
Other secondary outcomes	
KOOS Symptoms and Sports/Recreation subscores	± 8 points[1]
Isometric muscle strength	± ½ standard deviation of pooled baseline values
Dynamic assessment of pain	± 2.4
Pain self-efficacy questionnaire	± ½ standard deviation of pooled baseline values
EuroQoL EQ5D Questionnaire	± 0.32 points
Transition Questionnaire of global perceived change	± 3 points
[1] www.koos.nu accessed on October 23rd, 2021.	

9.3 Missing Data and Robustness

Our primary analyses will be based on the ITT population, including all randomised participants with available data at baseline. Missing data will be handled indirectly and statistically modelled using repeated-measures linear mixed models. These models will be valid if data are 'Missing at Random' (MAR): "Any systematic difference between the missing values and the observed values can be explained by differences in observed data" (15). Contrasts between groups will be estimated based on repeated-measures analysis of covariance applied in mixed linear models (i.e., at 12 and 26 weeks from baseline, respectively).

Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.

Loss to follow-up and missing data for various reasons is difficult to avoid in randomized trials and in particular in pragmatic trials. We will apply the analysis framework suggested by White et al. (2011) in which missing data related to the ITT approach depend on making plausible assumptions about the missingness of the data and including all participants in subsequent sensitivity analyses (16).

1. We attempt to follow up all randomized participants, even if they withdraw from allocated treatment (i.e., contact all individuals unless they explicitly stated that they had withdrawn their consent)
2. Perform a main analysis of all observed data that are valid under a plausible assumption about the missingness of the data (i.e., Model-based: data as observed; using repeated measures linear mixed models, assuming that data are 'Missing at Random' [MAR])
3. Perform sensitivity analyses to explore the effect of departures from the assumption made in the main (#2) analysis (i.e., a non-responder-imputation: using the value at baseline to replace missing data will correspond to a non-responder imputation; these models will potentially be informative even if data are 'Missing Not At Random' [MNAR])
4. Account for all randomized participants, at least in the sensitivity analyses (covered by #2 and #3 above), plus the corresponding analyses based on the PP population.

9.4 Primary analysis

Our primary analysis population will be all participants with available data at baseline, statistically modelled using repeated-measures linear mixed models (see above). These models will be valid if data are 'MAR'.

The primary analyses will be conducted according to the intention to treat principle. The ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen), rather than the actual treatment given (i.e., it is independent of treatment adherence). Accordingly, participants allocated to a treatment group will be followed up, assessed and analysed as members of that group, irrespective of their adherence to the planned course of treatment (i.e., independent of withdrawals and cross-over phenomena). Primary and secondary outcomes will be assessed using mixed linear models adjusted for baseline values.

9.4.1 Primary analysis of primary outcome

The primary outcome analysis will be an *equivalence analysis* based on the ITT population, asking whether the QE and HE treatments are equivalent regarding change from baseline in KUJALA questionnaire scores at the end of the treatment period (week 12). We will use a repeated measures linear mixed model regression analysis model adjusted for the baseline score of the KUJALA questionnaire. An interaction for time and group will be included.

$$KUJALA_{change} \approx GROUP + WEEK + GROUP \times WEEK + -KUJALA_{baseline}$$

Where GROUP has two levels (QE or. HE) and WEEK has three levels (0, 12, 26).

Analyses will include baseline and all follow-up data, and effects will be estimated at each follow-up visit; missing data will be handled implicitly via the mixed methods (maximum likelihood) approach. From this model the observed differences in the change from baseline in KUJALA questionnaire between QE and HE at week 12 will be estimated together with the associated 95% confidence interval (and the p-value) corresponding to the test of the hypothesis of no difference between treatments. The result of the primary analysis of the primary outcome will be presented in a table (shell table 2) and in a figure (shell Figure 2).

Equivalence will be claimed if the computed 95% confidence interval of the estimated group difference in the change from baseline in the KUJALA questionnaire at week 12 does not include ± 8 KUJALA points in the primary analysis.

Superiority will be claimed if the computed 95% confidence interval of the estimated group difference in the change from baseline in the KUJALA at week 12 does not include 0 in the primary analysis.

9.4.2 Primary analysis of secondary outcomes

The primary analyses of the key and other secondary outcomes will be *equivalence analyses* using the ITT population. Missing data will not be imputed.

The key secondary outcomes will be analysed identically to the primary outcome and adjusted the respective baseline value if available. We will compute differences with unadjusted two-sided 95% confidence intervals and p-values based on the equivalence paradigm. We will analyse the key

secondary outcomes in a prioritized order: The analyses of the secondary outcomes will be performed in sequence until one of the analyses fails to show equivalence.

The result of the primary analysis of the secondary outcomes for week 12 and week 26 will be presented in tables (shell Table 2 and 4).

9.5 Secondary analyses

First, we will repeat the primary analyses on the PP population that includes only participants who adhered to the allocated treatment without major protocol violations as defined above (section 8.3 and 8.4). These analyses will be conducted without imputation of missing data.

Secondly, we will adjust the primary analysis (on the ITT population without imputation of missing data) for a potential procedural mediator for the primary and key secondary outcomes: Satisfactory treatment adherence (in percentage):

$$VARIABLE_{change} \approx GROUP + WEEK + GROUP \times WEEK + VARIABLE_{baseline} + ADHERENCE$$

Where GROUP has two levels (QE or. HE) and WEEK has three levels (0, 12, 26).

Finally, we will perform an analysis of covariance of the primary and key secondary outcomes at week 12 (i.e. without the repeated measures) on the ITT population, with a baseline observation carried forward imputation of missing data at week 12 adjusted for the baseline values

$$VARIABLE_{change\_week12} = GROUP + VARIABLE_{baseline}$$

Where GROUP has two levels (QE or. HE).

If the sensitivity analyses are in agreement, and the sensitivity analyses and the primary analysis lead to essentially the same conclusions, confidence in the results is increased.

The result of the secondary analyses will be presented in supplementary tables (shell Appendix Tables S1-3).

9.6 Assessment of statistical assumptions

For the linear models of the primary and secondary outcomes, we will check for the normality of residuals by visual inspection of residual plots.

9.7 Statistical Software

The analysis will be carried out using the statistical software SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Linear mixed-effect models will be fitted using the MIXED procedure (proc mixed).

9.8 Harms

Analyses of adverse events (AEs) will be performed on the Safety Population (see section 8.5).

AEs will be categorised according to type of AE and assessed for relationship with the trial treatment and the number (and percentage) of related AE will be presented for each treatment arm. Deaths and AEs leading to discontinuation of study treatment will be listed. No formal statistical testing will be undertaken.

The AEs will be presented in a table (Shell Table 3)

9.9 Timing of analyses

When this statistical analysis plan was signed, recruitment to the COMPETE trial was completed (September 1, 2021), and the primary endpoint (12 weeks) had not been completed for all participants. We expect completion of the 12 weeks assessment for all participants by the beginning of December 2021 and completion of the 26 weeks assessment by the beginning of March 2022. We will close the database 2 months after the last participant's last assessment at the latest. Statistical analyses are expected to be completed after additionally 2 months at the latest.

10 DEVIATIONS FROM THE PROTOCOL

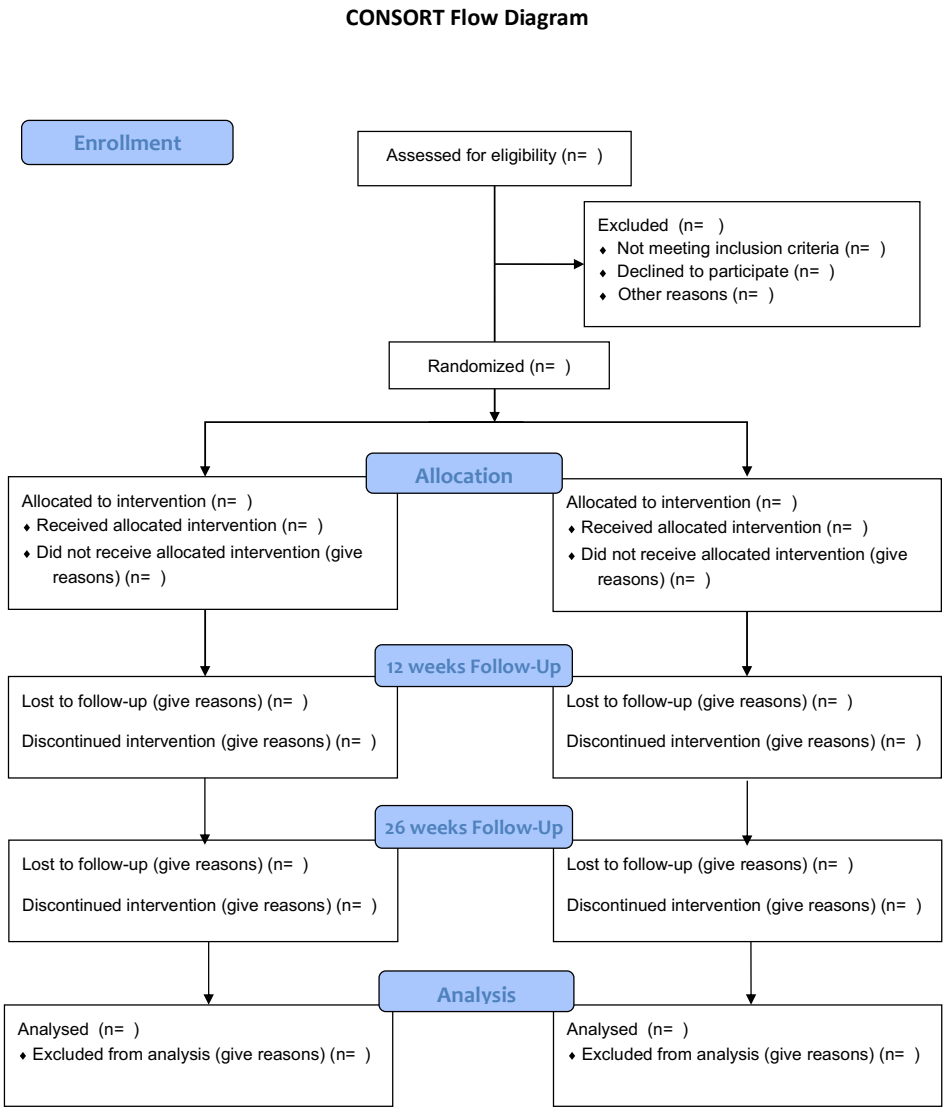
The following details in this SAP represents deviations from trial protocol version 1.5

Header in protocol	Change	Reason
9.2 Secondary outcomes	Secondary outcomes divided into 'Key Secondary Outcomes' and 'Other Secondary Outcomes'	To apply a hierarchy in the secondary outcomes.
9.2 Secondary outcomes Isometric muscle strength	Added hip adduction, flexion and internal rotation and knee flexion to the test battery	For a more comprehensive assessment of strength in the hip and knee region.

11 MANUSCRIPT OUTLINE

11.1 Shell Figure 1

Figure X: CONSORT flow diagram



11.2 Shell Table 1

Participants will be described with respect to baseline age, gender, height, body mass, Body Mass Index, and baseline values of primary and secondary outcomes if available, separately for the two randomised groups.

Continuous data will be summarised by mean, and SD. Categorical data will be summarised by numbers and percentages. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

Table X: Demographics and Baseline Characteristics

	Quadriceps exercise group (QE)	**Hip exercise group (HE)**
	n=	**n=**
Demographics		
Age, years	xx.x (xx.x)	xx.x (xx.x)
Female sex (n[%])	xx (xx.x%)	xx (xx.x%)
Body mass, kg	xx.x (xx.x)	xx.x (xx.x)
Height, m	xx.x (xx.x)	xx.x (xx.x)
Body Mass Index, BMI (kg/m2)	xx.x (xx.x)	xx.x (xx.x)
KUJALA questionnaire score (0-100)	xx.x (xx.x)	xx.x (xx.x)
KOOS (0-100)		
Pain	xx.x (xx.x)	xx.x (xx.x)
Physical Function	xx.x (xx.x)	xx.x (xx.x)
Symptoms	xx.x (xx.x)	xx.x (xx.x)
QoL	xx.x (xx.x)	xx.x (xx.x)
Sports & Recreation	xx.x (xx.x)	xx.x (xx.x)
Dynamic assessment of pain (0-10)	xx.x (xx.x)	xx.x (xx.x)
Isometric muscle strength		
Hip abductors (N)	xx.x (xx.x)	xx.x (xx.x)
Hip adductors (N)	xx.x (xx.x)	xx.x (xx.x)
Hip extensors (N)	xx.x (xx.x)	xx.x (xx.x)
Hip flexors (N)	xx.x (xx.x)	xx.x (xx.x)
Hip external rotators (N)	xx.x (xx.x)	xx.x (xx.x)
Hip internal rotators (N)	xx.x (xx.x)	xx.x (xx.x)
Knee extensors (quadriceps) (N)	xx.x (xx.x)	xx.x (xx.x)
Knee flexors (hamstrings) (N)	xx.x (xx.x)	xx.x (xx.x)
Pain Self-efficacy questionnaire (0-60)	xx.x (xx.x)	xx.x (xx.x)
EuroQoL EQ5D Questionnaire (-0.624 to 1.000)	xx.x (xx.x)	xx.x (xx.x)

11.3 Shell Table 2

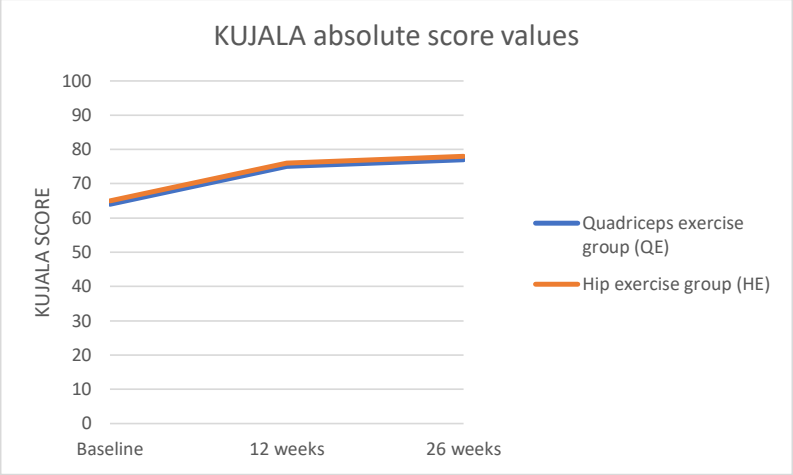
Table X: Primary and Secondary Outcomes at week 12 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=)	HE (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Change in KUJALA questionnaire – score (0 to 100); equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Change in KUJALA questionnaire – score (0 to 100); superiority test*				0.xxx*
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in isometric muscle strength				
Hip abductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip adductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip extensors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip flexors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip external rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip internal rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee extensors (quadriceps) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee flexors (hamstrings) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in Dynamic Assessment of Pain (VRS (0-10))	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in EQ5D Questionnaire (index -0.624 to 1,000)	x.xxx (x.xxx)	x.xxx (x.xxx)	x.xxx (x.xxx to x.xxx)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	x.x (xx.x)	x.x (x.x)	x.x (x.x to x.x)	
Treatment adherence				
Treatment adherence (%)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Treatment adherence ≥66% - no. (%)	xx (xx.x %)	xx (xx.x%)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
*Primary outcome will be analysed using both a test for equivalence and a test for superiority.
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

11.4 Shell Figure 2

Figure X: Exemplar (hypothetical) trajectories for our primary efficacy outcome measure (i.e. absolute score values in KUJALA questionnaire) in the ITT population.



11.5 Shell Table 3

Table X. Adverse events in the intention-to-treat population.

	QE (n=)	HE (n=)
Exposure time – patient weeks		
AE - no. of patients (%)		
AE - no. of events (rate – event per patient week)		
AEs leading to discontinuation - no. of patients (%)		
AEs, relationship to trial treatment, no. of events (rate – event per patient week)		
Not related		
Probably not related		
Probably related		
AEs, classification, no. of events (rate – event per patient week)		
PFP pain exacerbation		
Muscle soreness		
Other		
Deaths - no. of events (rate – event per patient week)		
AE; Adverse event. The severity of an adverse event refers to the maximum intensity of the event. An event was considered severe (compared with mild or moderate) if it interfered substantially with the patient's usual activities.		

11.6 Shell Table 4

Table X: Primary and Secondary Outcomes at week 26 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=)	HE (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Change in KUJALA questionnaire – score (0 to 100); equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Change in KUJALA questionnaire – score (0 to 100); superiority test*				0.xxx*
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Change in EQ5D Questionnaire (index 1,000 to -0.624)	x.xxx (x.xxx)	x.xxx (x.xxx)	x.xxx (x.xxx to x.xxx)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	x.x (xx.x)	x.x (x.x)	x.x (x.x to x.x)	
Values are least squares means ± standard error unless otherwise stated. *Primary outcome will be analysed using both a test for equivalence and a test for superiority. KOOS: Knee injury and osteoarthritis outcome score. VRS: Verbal Rating Scale				

11.7 Shell Appendix Table S1

Table X: Primary and Key Secondary Outcomes at week 12 and 26 in the PP population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=)	HE (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Week 12: Change in Kujala questionnaire – score (0 to 100); equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Week 12: Change in Kujala questionnaire – score (0 to 100); superiority test*				0.xxx*
Week 26: Change in Kujala questionnaire – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Key Secondary outcomes:				
Week 12: Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary outcomes	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in isometric muscle strength				
Hip abductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip adductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip extensors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip flexors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip external rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip internal rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee extensors (quadriceps) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee flexors (hamstrings) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in Dynamic Assessment of Pain (VRS (0-10))	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in EQ5D Questionnaire (index 1,000 to -0.624)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in EQ5D Questionnaire (index 1,000 to -0.624)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Values are least squares means ± standard error unless otherwise stated. *Primary outcome will be analysed using both a test for equivalence and a test for superiority. KOOS: Knee injury and osteoarthritis outcome score. VRS: Verbal Rating Scale				

11.8 Shell Appendix Table S2

Table X: Primary and Key Secondary Outcomes at week 12 and 26 in the ITT population adjusted for treatment adherence. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	QE (N=)	HE (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Week 12: Change in Kujala questionnaire – score (0 to 100); equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Week 12: Change in Kujala questionnaire – score (0 to 100); superiority test*				0.xxx*
Week 26: Change in Kujala questionnaire – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Key Secondary outcomes:				
Week 12: Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Other Secondary outcomes	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in isometric muscle strength				
Hip abductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip adductors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip extensors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip flexors (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip external rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Hip internal rotators (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee extensors (quadriceps) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Knee flexors (hamstrings) (N)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in Dynamic Assessment of Pain (VRS (0-10))	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in EQ5D Questionnaire (index 1,000 to -0.624)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Sports and recreation– score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in KOOS Symptoms – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Change in EQ5D Questionnaire (index 1,000 to -0.624)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 26: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	

Values are least squares means ± standard error unless otherwise stated.
*Primary outcome will be analysed using both a test for equivalence and a test for superiority.
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

11.9 Shell Appendix Table S3

Table X: Primary and Key Secondary Outcomes at week 12 in the ITT population. CI denotes 95% confidence interval. Based on analysis of covariance, where missing data is conservatively imputed using baseline observation carried forward.

	QE (N=)	HE (N=)	Estimated treatment difference	P-value
	Mean (SE)	Mean (SE)	Mean (95% CI)	
Primary outcome:				
Week 12: Change in Kujala questionnaire – score (0 to 100); equivalence test*	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	0.xxx*
Week 12: Change in Kujala questionnaire – score (0 to 100); superiority test*				0.xxx*
Key Secondary outcome:				
Week 12: Change in KOOS Pain – score (0-100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Function – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Week 12: Change in KOOS Quality of life – score (0 to 100)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x to xx.x)	
Values are least squares means ± standard error unless otherwise stated. *Primary outcome will be analysed using both a test for equivalence and a test for superiority. KOOS: Knee injury and osteoarthritis outcome score.				

12 REFERENCES

1. Dolak KL, Silkman C, Medina McKeon J, Hosey RG, Lattermann C, Uhl TL. Hip strengthening prior to functional exercises reduces pain sooner than quadriceps strengthening in females with patellofemoral pain syndrome: a randomized clinical trial. J Orthop Sports Phys Ther. 2011;41(8):560-70.
2. Khayambashi K, Fallah A, Movahedi A, Bagwell J, Powers C. Posterolateral hip muscle strengthening versus quadriceps strengthening for patellofemoral pain: a comparative control trial. Arch Phys Med Rehabil. 2014;95(5):900-7.
3. Ferber R, Bolgla L, Earl-Boehm JE, Emery C, Hamstra-Wright K. Strengthening of the hip and core versus knee muscles for the treatment of patellofemoral pain: a multicenter randomized controlled trial. J Athl Train. 2015;50(4):366-77.
4. Saad MC, Vasconcelos RA, Mancinelli LVO, Munno MSB, Liporaci RF, Grossi DB. Is hip strengthening the best treatment option for females with patellofemoral pain? A randomized controlled trial of three different types of exercises. Braz J Phys Ther. 2018;22(5):408-16.
5. Hott A, Brox JI, Pripp AH, Juel NG, Paulsen G, Liavaag S. Effectiveness of Isolated Hip Exercise, Knee Exercise, or Free Physical Activity for Patellofemoral Pain: A Randomized Controlled Trial. Am J Sports Med. 2019;47(6):1312-22.
6. Esculier JF, Roy JS, Bouyer LJ. Psychometric evidence of self-reported questionnaires for patellofemoral pain syndrome: a systematic review. Disabil Rehabil. 2013;35(26):2181-90.
7. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. Scand J Med Sci Sports. 2010;20(3):493-501.
8. Romero-Franco N, Jimenez-Reyes P, Montano-Munuera JA. Validity and reliability of a low-cost digital dynamometer for measuring isometric strength of lower limb. J Sports Sci. 2017;35(22):2179-84.
9. Mentiplay BF, Perraton LG, Bower KJ, Adair B, Pua YH, Williams GP, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. PLoS One. 2015;10(10):e0140822.
10. Klokker L, Christensen R, Waehrens EE, Bandak E, Bartholdy C, Bliddal H, et al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: construct validity, responsiveness, and interpretability in a research setting. Health Qual Life Outcomes. 2016;14:91.
11. Chiarotto A, Vanti C, Cedraschi C, Ferrari S, de Lima ESRF, Ostelo RW, et al. Responsiveness and Minimal Important Change of the Pain Self-Efficacy Questionnaire and Short Forms in Patients With Chronic Low Back Pain. J Pain. 2016;17(6):707-18.
12. Vejlgaard C, Maribo T, Riisgaard Laursen J, Schmidt AM. Reliability and smallest detectable change of the Danish version of the Pain Self-Efficacy Questionnaire in patients with chronic low back pain. Scand J Pain. 2021;21(4):809-13.
13. Bilbao A, Garcia-Perez L, Arenaza JC, Garcia I, Ariza-Cardiel G, Trujillo-Martin E, et al. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. Qual Life Res. 2018;27(11):2897-908.
14. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-70.
15. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. JAMA. 2016;315(4):407-8.
16. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011;342:d40.

Appendix 5 – Statistical Analysis Plan (SAP) Study 2

Version 1
Date: 19-11-2021

Which patient characteristics can predict response to quadriceps or hip muscles based resistance training programs: Statistical Analysis Plan for Secondary Analyses of the COMPETE Trial

Authors:
Rudi Hansen, Marius Henriksen

Collaborators/Investigators:
Michael Skovdal Rathleff, Niels Christoffer Brushøj

1 PURPOSE

The purpose of this statistical analysis plan (SAP) is to describe detailed aspects of data preparation and analyses. This SAP was set up before starting the analyses and is based on the final trial protocol (version 1.2, April 18, 2017).

2 INTRODUCTION

Patellofemoral Pain (PFP) is a common knee problem, which particularly affects adolescents and young adults (1-4). PFP is characterised by significant retropatellar and/or peripatellar pain and impairment of function and quality of daily life (5). Exercise has repeatedly been shown beneficial for pain and physical function and is unequivocally recommended as a core component of the management of PFP (6, 7). Different types of exercise (e.g. quadriceps strengthening, hip strengthening and functional/neuromuscular exercises) have been investigated. In general, these different types of exercises produce similar small to moderate beneficial effects in pain and physical function (8). However, the PFP population is very heterogeneous and "one-size-fits-all"-approaches presumably are sub-optimal because the heterogeneity is ignored. This heterogeneity probably explains the overall limited beneficial effects of exercise, and the lack of differences in direct comparisons of different exercise types.

Extensive effort has been made to identify indicators of prediction for conservative PFP management. In a recent systematic review, Mathews et al. aimed to determine which baseline patient characteristics, that were associated with patient outcome from a specific treatment (9). It included 24 studies evaluating 180 participant characteristics and found 16 factors that were associated with a poor outcome, with longer duration of symptoms being the most reported. Two studies in the review investigated predictors of successful outcome after exercise and they identified only patellar tilt angle difference, which is the difference in patellar tilt angle in a quadriceps contracted and a quadriceps relaxed position measured on axial CT, to be associated with a successful outcome of a leg press strengthening and lower limb stretching programme. Within the studies identified in the systematic reviews, the low number of participants per predictor variable, and the absence of comparator groups, may affect the validity of the predictors. In general, methodological limitations make it difficult to predict the outcome after one specific treatment compared with another, and adequately designed randomised trials are warranted to identify treatment effect modifiers (9).

A useful method for identifying treatment modifiers is through secondary analyses of RCT data. Effect modifiers are patient characteristics measured at baseline, that interact with the treatment to influence clinical outcomes, such that those with different levels of the moderator respond differently to the treatment compared to the alternate treatment (10). Identifying moderators of outcomes may, therefore, help clinicians and researchers to direct treatments to patient subgroups most likely to benefit. Recently such a post-hoc analysis has been published, with

the aim to identify predictors of outcomes for function, pain, and measures of change after 1 year for patients with PFP who participated in a randomized controlled trial (11). As there were no significant between-group difference in the investigated interventions, the data were analyzed as a single cohort. It revealed that number of pain sites at baseline was most consistent predictor of outcome at 1 year. Baseline values of Anterior Knee pain Scale (AKPS) and worst pain were significant predictors of outcome of the same variable at 1 year. Low self-efficacy, longer pain duration, and male sex were also significant predictors of poorer outcomes at 1 year. More RCTs with evaluation of outcome prediction as a primary aim are clearly warranted to provide clinicians with robust evidence and facilitate evidence-informed, tailored intervention to this heterogeneous patient population (12).

We are conducting a randomized controlled study comparing the efficacy of quadriceps and hip muscles-based resistance training programs with a prespecified aim to explore candidate patient characteristics that predict differential responses to the two exercise programs on self-reported pain and physical function in individuals with PFP. Based on the RCT, the purpose of this study is to identify clinically feasible outcome predictors of comparative treatment effect of either protocol making it possible for clinicians to target the intervention in the future. To our knowledge no randomized controlled trial has been conducted with the aim to identify outcome predictors.

3 AIM AND OBJECTIVES

Stratified medicine refers to the targeting of treatments (including pharmacological and non-pharmacological interventions) according to the biological or risk characteristics shared by subgroups of patients. Thus, stratified medicine uses baseline information about a patient's likely response to treatment to tailor treatment decisions. The aim of the secondary analyses will be to identify contextual factors that might modify the observed treatment effect (Quadriceps Exercise program (QE) *vs* Hip muscle Exercise program (HE)) across patient subgroups, i.e. whether the treatment effect is modified by the value of a variable assessed at baseline. Our objective is to explore a range of baseline factors that could potentially modify any difference in the effects on pain and function between QE and HE assessed by the change from baseline in the KUJALA questionnaire score after 12 weeks.

4 STUDY DESIGN

4.1 Trial Design

The trial is a single centre, randomized, parallel-group, 26 weeks (6 months), equivalence trial comparing a 12-weeks focused "Quadriceps Exercise" protocol and a 12-weeks focused "Hip Exercise" protocol with a primary endpoint at 12 weeks (after treatment) and a follow-up at 26 weeks. The trial is conducted among patients with PFP. A total of 200 patients has been randomly assigned on a 1:1 basis to one of the two treatments, QE or HE.

4.2 Analysis of effectiveness according to patient subgroups

The analyses of effectiveness are based on assumptions that treatment effects vary according to an interaction between group allocation and certain patient characteristics that can be binary defined as present/absent. Continuous data is dichotomised whenever a reasonable and clinically relevant threshold can be set.

The analyses will be done on the ITT population as defined in section 6.1.

The analyses will focus on the interaction between presence/absence of a certain characteristic and group allocation (QE vs. HE) at each time point (week 12 and 26). Changes from baseline in outcomes will be analysed using repeated measures mixed linear models with group (QE vs HE), week (12 and 26), and candidate baseline predictor characteristic (present/absent) as fixed factors,

and participant as random factor. The models will be adjusted for the outcome baseline value. We will focus on the triple interaction part of the model (group×week×predictor) that will be broken down by pair-wise comparisons of group and predictors at each week. All statistical tests will be 2-tailed at an alpha = 0.05 with estimates presented with 95% confidence intervals.

5 OUTCOMES

5.1 Primary outcome

The primary outcome is assessed at week 12 as change from baseline in the KUJALA questionnaire – a widely used and well-validated survey instrument evaluating pain and function in PFP. We will measure the difference in the changes from baseline in the KUJALA score in the study knee between QE vs HE after 12 weeks.

5.2 Potential effect modifiers

The candidate baseline characteristics encompass self-reported information as well as clinical observations and tests. We have a priori identified the following baseline variables that we wish to explore as potential effect modifiers, based on findings in previous studies and clinical experience. (Table A):

Table A: Dichotomous or dichotomised baseline variables

Variables	Description	Values (categories)
Low back pain	Self-reported presence of low back pain during the last 3 months, recorded as "Yes" (pain present) or "No" (No pain). Predefined frequency options are available if answered "Yes": Almost daily, Several times during a week, weekly, monthly, rarely. In a post hoc analysis of an RCT, higher number of pain sites throughout the body is found to be a predictor of poor outcome and less change at 1 year (13). The rationale, therefore, counts for the subsequent variables (presence of bilateral pain, presence of ankle pain, and presence of hip pain). We will a priori define a dichotomisation of the scores "Almost daily", "Several times during a week", "Weekly", "Monthly" as "Low back pain present" and the scores "No" and "Rarely" as "Low back pain not present".	0 = Low back pain not present 1 = Low back pain present
Presence of bilateral knee pain	Self-reported presence of pain in the contralateral knee (not target knee) during the last 3 months, recorded as "Yes" (pain present) or "No" (No pain). Predefined frequency options are available if answered "Yes": Almost daily, Several times during a week, weekly, monthly, rarely. We will a priori define a dichotomisation of the scores "Almost daily", "Several times during a week", "Weekly", "Monthly" as "Pain in the contralateral knee present" and the scores "No" and "Rarely" as "Pain in the contralateral knee not present".	0 = Pain in the contralateral knee not present 1 = Pain in the contralateral knee present
Presence of ankle pain	Self-reported presence of ankle pain (one or both ankles) during the last 3 months, recorded as "Yes" (pain present) or "No" (No pain). Predefined frequency options are available: Almost daily, Several times during a week, weekly, monthly, rarely We will a priori define a dichotomisation of the scores "Almost daily", "Several times during a week", "Weekly", and "Monthly" as "Ankle pain present" and the scores "No" and "Rarely" as "Ankle pain not present".	0 = Ankle pain not present 1 = Ankle pain present
Presence of hip pain	Self-reported presence of hip pain (one or both hips) during the last 3 months, recorded as "Yes" (pain present) or "No" (No pain). Predefined frequency options are available: Almost daily, Several times during a week, weekly, monthly, rarely We will a priori define a dichotomisation of the scores "Almost daily", "Several times during a week", "Weekly", and "Monthly" as "Hip pain present" and the scores "No" and "Rarely" as "Hip pain not present".	0 = Hip pain not present 1 = Hip pain present

Variables	Description	Values (categories)
Body mass index (BMI)	The participants' BMI is measured at baseline and dichotomized with a cut-off at 25 (BMI ≥ 25). BMI is used to broadly define different weight groups in adults, and a BMI at or above 25 is defined as overweight (14).	0 = BMI< 25 1 = BMI≥ 25
Sex	Female or male sex	0 = female 1 = male
Duration of knee pain	Generally, pain is regarded as chronic when it lasts or recurs for more than 3 to 6 months (15). Longer duration of pain is the most reported factor of outcome success in a systematic review on predictors (12). We a priori define chronic pain as pain lasting for more than 6 months.	0 = Knee pain not chronic 1 = Knee pain chronic
Education level	Self-reported highest level of education: "Primary school", "Craftsman", "Highscool", "Short higher education (< 3 years), "Medium-term higher education (3-4 years)", and "Longer higher education (>4 years)". No rationale for this variable is found in the literature, and it is considered purely explorative. We will a priori define a dichotomization of the scores "Medium-term higher education (3-4 years)", and "Longer higher education (>4 years)" as "Long education" and the scores "Primary school", "Craftsman", "Highscool", "Short higher education (< 3 years) as "Short education".	0 = Long education 1 = Short education
Occupation	Participants are asked about their status concerning occupation and education. Answer categories are "Currently not working", "Currently studying", and "Currently working". No rationale for this variable is found in the literature, and it is considered purely explorative. We will a priori define a dichotomization of the scores "Currently studying" and "Currently working" as "Currently studying/working" and the score "Currently not working" as "Currently not working/not studying".	0 = Currently working/studying 1 = Currently not working/not studying
Hypermobility	Hypermobility is assessed by the Beighton Score applying the revised criteria for the diagnosis of benign joint hypermobility syndrome. The Beighton score ranges from 0-9. Generalized joint hypermobility has been linked to the development of PFP (16), In this study we define hypermobility as a score of 4 or more, which is generally considered an indication of joint hypermobility (17, 18).	0 = Normal joint mobility 1 = Generalized joint hypermobility
Knee joint valgus malalignment during a forward lunge movement	Knee joint alignment is assessed by clinically observation of the participant while he/she performs a forward lunge movement. The observation is made by a trained physiotherapist that classifies each participant's knee as 'definite valgus present', 'No evidence of dynamic malalignment and 'definite varus present' (19). Valgus malalignment is generally considered a predisposing factor for patellofemoral pain (20, 21). We will a priori define a dichotomisation of the score "definite valgus present" as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".	0 = No valgus malalignment 1 = Valgus malalignment
Knee joint valgus malalignment during a single-leg squat movement	Knee joint alignment is assessed by clinical observation of the participant while he/she performs a single-leg squat. The observation is made by a trained physiotherapist that classifies each participant's knee as 'definite valgus present', 'No evidence of dynamic malalignment and 'definite varus present' (19). Valgus malalignment is generally considered a predisposing factor for patellofemoral pain (20, 21). We will a priori define a dichotomisation of the score "definite valgus present" as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".	0 = No valgus malalignment 1 = Valgus malalignment
Exercise self-efficacy	Exercise self-efficacy in relation to the two different exercise programs is assessed by asking the participants to rate their confidence in performing the allocated exercise program on an 11-point (0-10) Likert scale with 0 representing "Not at all confident" and 10 representing "Completely confident". In a recent RCT, higher self-efficacy as measured by the Knee Self-Efficacy Scale, which pertain to how certain respondents feel about performing various activities, were predictors of global change after an exercise intervention (11).	0 = High self-efficacy 1 = Low self-efficacy

Variables	Description	Values (categories)
	We will a priori define a dichotomisation of the scores scores 6-10 as "High self-efficacy" and the scores 0-5 as "Low self-efficacy".	
Neuropatic pain	The painDETECT questionnaire (PDQ) is a patient reported questionnaire developed and validated to assess presence of signs of neuropathic pain. It comprises questions regarding pain intensity, course of pain, subjective experience of a radiating quality of the pain, and the presence and perceived severity of seven somatosensory symptoms of neuropathic pain. A validated algorithm is used to calculate a total score ranging from -1 to 38. A score ≤12 indicates that the presence of neuropathic pain is unlikely, a score ≥19 indicates that a neuropathic pain component is likely to be present, while a score of 13–18 points towards unclear screening conclusion (22). Studies indicate decreased pressure pain thresholds, increased tactile detection thresholds, and increased warmth detection, all indicative of a neuropatic pain component, in PFP patients compared to controls (23, 24) We define a dichotomization of the scores as "Neuropathic pain component" (scores ≥19) and "No neuropathic pain component (scores ≤18).	0 = No neuropathic pain component 1 = Neuropathic pain component
Pain severity	As a part of the painDETECT questionnaire, participants are asked to rate their average pain during the past 4 weeks on a 0-10 Numeric Rating Scale. Results from the literature show that the cut-off point between mild and moderate pain, in terms of pain related interference with functioning, is placed between 2.5 and 4.5 and the cut point between moderate and severe pain is between 5.5 and 7.4 (25). Accordingly, we define a dichotomization of the scores 0-6 as "Mild or moderate pain" and the scores 7-10 as "Severe pain". Usual pain intensity has been identified as possible outcome predictors for a taping intervention, foot orthoses and for a tailored strengthening, stretching, and mobilisation programme, but the evidence is considered limited (12).	0 = Mild or moderate pain 1 = Severe pain
Pain Catastrophizing Scale	The Pain Catastrophizing Scale (PCS) will be used as a measure of pain-related catastrophic thinking. The PCS instructs participants to reflect upon past painful experiences, and to indicate the frequency with which 13 pre-specified thoughts or feelings occur while they are experiencing pain. The frequency is scored on a 5-point scale ranging from 0 (not at all) to 4 (all the time). The PCS measures 3 distinct components: rumination, magnification, and helplessness. Higher levels of psychological impairment, including catastrophizing thoughts has been found in PFP patients with more-severe PFP-related disability than less-severe cases (26). Also, changes in catastrophizing thoughts after treatment, has been shown to predict the changes in both pain and disability in patients suffering from patellofemoral pain (27). Previous studies have shown a cut-off of more than 30 points to be associated with clinical relevance (28-30). Accordingly, we define a dichotomization of the scores as "Pain catastrophizing" (scores > 30) and "No pain catastrophizing" (scores 0 - 30).	0 = No pain catastrophizing 1 = Pain catastrophizing
Pain self-efficacy	The pain self-efficacy questionnaire is a 10-item questionnaire developed to assess the confidence people with pain have in performing activities while in pain (25). It is applicable to all persisting pain presentations and covers a range of functions. Confidence in performing activities is rated on a 7-point (0-6) Likert scale with 0 representing not at all confident and 6 representing completely confident. A total score is calculated by summing the answers producing a score between 0 and 60. Higher scores reflect stronger self-efficacy beliefs. Among injured workers, raw scores around 40 (percentile = 50) are associated with return to work and maintenance of functional gains, whilst lower scores tend to predict less sustainable gains (31). We define a dichotomization of the scores scores 0-39 as "Poor pain self-efficacy and the scores 40-60 as "Good self-efficacy".	0 = Good pain self-efficacy 1 = Poor pain self-efficacy
Midfoot mobility magnitude	Midfoot mobility is calculated as the change in midfoot width from non-weight bearing to weight bearing. Midfoot width is measured at 50% of total foot length using a calliper during non-weight bearing and during standing with equal weight on each foot. The mobility is recorded as the difference between non-weight bearing and weight bearing measurements.	0 = Normal or limited midfoot mobility magnitude 1 = Excessive mobility magnitude

Variables	Description	Values (categories)
	Normative data for the foot mobility magnitude has established by assessment of 345 healthy adult participants. The mean change in midfoot width between weight bearing and non-weight bearing was 0.92 cm (SD 0.32 cm) in the left foot and 0.88 cm (SD 0.33) in the right foot for females. For males, the mean change was 1.02 cm (SD 0.34) in the left foot and 1.00 cm (SD 0.32 cm) in the right foot for males (32). A more pronated foot posture and greater midfoot mobility is associated with worse knee pain and greater disability in individuals with patellofemoral osteoarthritis (33). A recent RCT investigated the effect of foot orthoses vs. hip exercises and found no association between midfoot width mobility and treatment outcome in patients with patellofemoral pain (34). We define a dichotomization of scores within the range 0-1.24 cm (0.92 cm +1*SD) as "Normal or limited midfoot mobility magnitude" and the scores above 1.25 cm as "Excessive midfoot mobility magnitude" for the females. For the males, we define a dichotomization of scores within the range 0-1.36 cm (1.02 cm +1*SD) as "Normal or limited midfoot mobility magnitude and the scores above 1.37 cm as "Excessive midfoot mobility magnitude". Excessive midfoot mobility has been linked to the development of PFP (35).	
Age	Age at inclusion in the trial. To split the trial population in two groups, we chose to categorize the participants based on the median age of all trial participants in the ITT population.	0 = Below median age 1 = Above median age

6 Study populations

6.1 Intention-To-Treat population

The Intention-To-Treat (ITT) population consist of all randomized patients irrespective of whether the patient actually received study intervention or the patient's compliance with the study protocol, in the treatment group to which the participant was assigned at randomisation (Intention-To-Treat principle). A patient will be considered randomised as soon as a treatment is assigned according to the allocation sequence.

The participant demographics and baseline data for the ITT population will be summarized in a table (shell Table 1). Participants will be described with respect to baseline age, sex, body mass, height, body mass index, duration of pain, primary outcome and the prespecified potential effect modifiers, separately for the two groups.

Continuous data will be summarized by means and SDs (or medians and interquartiles, depending on the data distribution). Categorical data will be summarized by numbers and percentages. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted based on the calculated standardized differences.

6.2 Modified Intention-To-Treat population

The modified Intention-To-Treat (mITT) population consist of patients in the ITT population (see above) with recorded baseline data on the potential effect modifier variable under analysis.

6.3 As Observed population

The As Observed (AO) population consists of all participants for whom baseline data and primary outcome data have been observed and no imputation will be carried out.

7 Data preparation

7.1 Changes from baseline

The primary outcome is change from baseline in KUJALA at week 12 This will be calculated for each individual as the baseline value subtracted from the week 12 value:

$$KUJALA_{change} = KUJALA_{week12} - KUJALA_{baseline}$$

Thus, a positive change value indicates that the week 12 value is greater than the baseline value, which suggest an improvement in the KUJALA (i.e. less pain and improved function).

8 STATISTICAL ANALYSES

8.1 General considerations

In the primary analysis, all participants will be analysed using the ITT population (or mITT) according to the intention-to-treat principle. We will use a repeated measures linear mixed model regression analysis model adjusted for the baseline score of the KUJALA. An interaction for time, week, and group will be included. All 95% confidence intervals and *P*-values will be two sided. We will not apply explicit adjustments for multiplicity, rather we will explicitly state that the results are exploratory and hypothesis generating.

8.2 Missing Data and Robustness

Our primary analyses will be based on the ITT population, including all randomized participants with available data at baseline. Missing data will be handled implicitly via the mixed methods (maximum likelihood) approach that are valid if missing data are missing at random (MAR).

Missing data on the prespecified potential effect modifiers measured at baseline will not be imputed. Rather, we will explicitly report any such missing data and the primary analyses will be done on the modified intention-to-treat population (mITT) for that potential baseline effect modifier.

To assess the robustness of the primary analyses, we will repeat it on the ITT population with missing data on the primary outcome variable at week 12 and 26 conservatively replaced with the baseline observation (non-responder imputation). Also, the primary analysis will be repeated on the AO population with no imputation of missing data. The results of these sensitivity analyses will be presented in appendix tables.

8.3 Primary analyses

We will use linear regression analysis models.
For the analyses of whether the dichotomous baseline variables modify the treatment effect (primary outcome; change from baseline in KUJALA score) the following repeated measures linear mixed model is used:

$$\begin{aligned} KUJALA_{change} \approx\ & GROUP + MODIFIER + WEEK + \\ & GROUP \times MODIFIER + \\ & GROUP \times WEEK + \\ & MODIFIER \times WEEK + \\ & GROUP \times MODIFIER \times WEEK + KUJALA_{baseline} \end{aligned}$$

Where Group, Week, and "Modifier Covariate" will be included as a Fixed Effect Class variable and the participant included as Random Effect. Analyses will include all collected data at all time points; missing data will be handled implicitly via the mixed methods (maximum likelihood) approach. From this model, the observed differences in QE and HE at week 12 and 26 in the change from baseline in KUJALA score between subgroups of participants based on the presence of the potential effect modifiers will be estimated together with the associated 95% confidence interval (and the *P*-value) corresponding to the test of the hypothesis of no interaction between group and treatment modifier. The result will be presented in a figure (shell Figure 2).

8.4 DEVIATIONS FROM THE PROTOCOL

The following details in this SAP represents deviations from trial protocol version 1.5

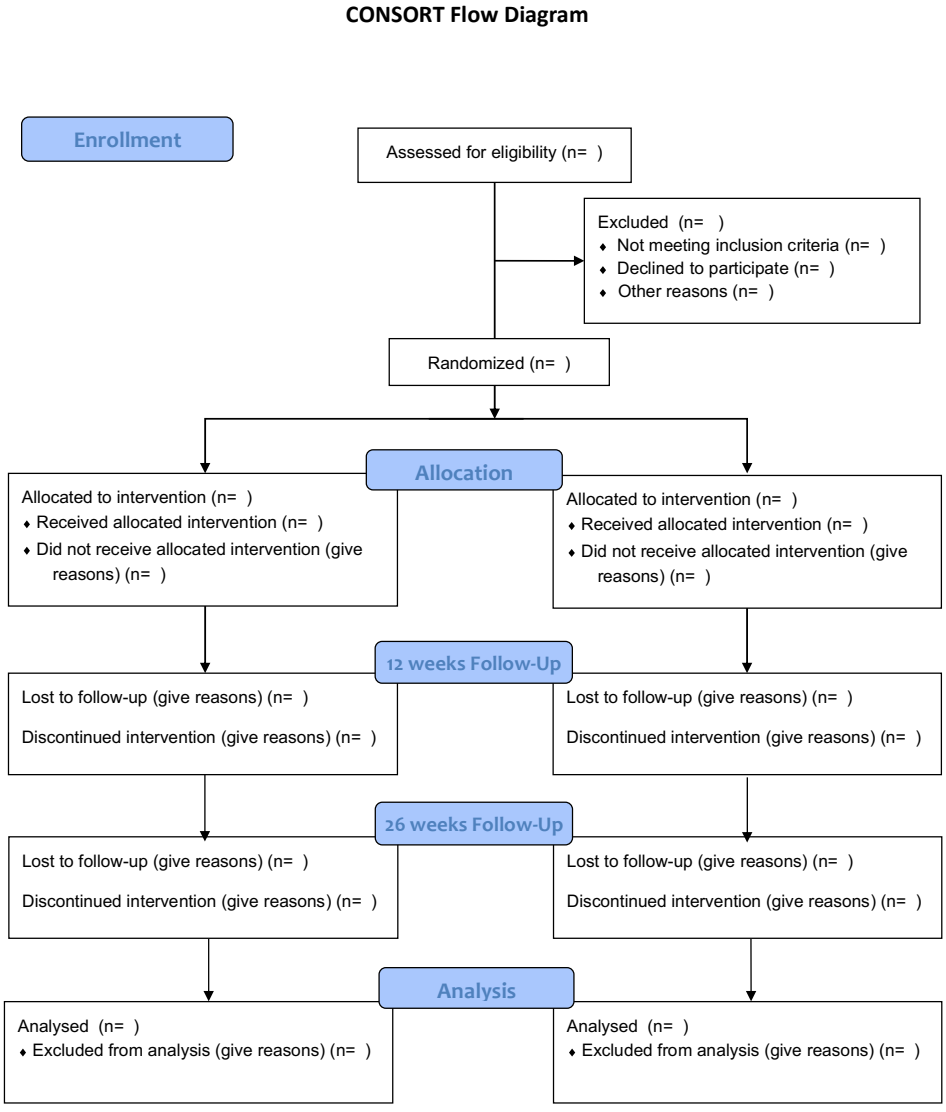
Header in protocol	**Change**	**Reason**
15.6 Analysis of effectiveness according to patient subgroups	In the protocol it was stated that the study population would be split in two: an exploration cohort and a replication cohort in order to replicate our results immediately (i.e. without running another study). This analytical approach has been abandoned and the study population will be analysed as one.	Subgroup analyses of RCT are prone to loss of statistical power with a subsequent increased risk of both type I and type II errors. To reduce this risk, we have chosen not to split the study population in two.
15.6 Analysis of effectiveness according to patient subgroups	In the protocol it was stated that the subgroup analyses should be done on the as-observed population. This has been changed to the ITT population.	ITT analyses are most appropriate in analyses of superiority.
5.2 Potential effect modifiers	The baseline characteristic "Physiotherapist estimated prognosis for the participant" is deleted from study.	Physiotherapist prognosis is not a potential effect modifier and will be analysed in a separate manuscript.

9 Manuscript outline

9.1 Shell Table 1. Baseline characteristics and summary of treatment effect modifiers

	Quadriceps exercise group (QE)	Hip exercise group (HE)
	n=	n=
Demographics		
Age, years	xx.x (xx.x)	xx.x (xx.x)
Duration of pain, months	xx.x (xx.x)	xx.x (xx.x)
Body mass, kg	xx.x (xx.x)	xx.x (xx.x)
Height, m	xx.x (xx.x)	xx.x (xx.x)
Body Mass Index, BMI (kg/m2)	xx.x (xx.x)	xx.x (xx.x)
Primary outcome		
KUJALA questionnaire score (0-100)	xx.x (xx.x)	xx.x (xx.x)
Potential effect modifiers		
Female sex, n(%)	xx (xx.x%)	xx (xx.x%)
Above median age, n(%)	xx (xx.x%)	xx (xx.x%)
BMI≥ 25, n(%)	xx (xx.x%)	xx (xx.x%)
Chronic Knee pain (>6 months), n(%)	xx (xx.x%)	xx (xx.x%)
Presence of low back pain, n(%)	xx (xx.x%)	xx (xx.x%)
Presence of bilateral knee pain, n(%)	xx (xx.x%)	xx (xx.x%)
Presence of ankle pain , n(%)	xx (xx.x%)	xx (xx.x%)
Presence of hip pain, n(%)	xx (xx.x%)	xx (xx.x%)
Short education, n(%)	xx (xx.x%)	xx (xx.x%)
Currently not working/not studying, n(%)	xx (xx.x%)	xx (xx.x%)
Generalized joint hypermobility, n(%)	xx (xx.x%)	xx (xx.x%)
Knee joint valgus malalignment forward lunge, n(%)	xx (xx.x%)	xx (xx.x%)
Knee joint valgus malalignment single-leg squat, n(%)	xx (xx.x%)	xx (xx.x%)
Low self-efficacy, n(%)	xx (xx.x%)	xx (xx.x%)
Neuropathic pain component, n(%)	xx (xx.x%)	xx (xx.x%)
Severe pain, n(%)	xx (xx.x%)	xx (xx.x%)
Pain catastrophizing, n(%)	xx (xx.x%)	xx (xx.x%)
Poor pain self-efficacy, n(%)	xx (xx.x%)	xx (xx.x%)
Excessive mobility magnitude, n(%)	xx (xx.x%)	xx (xx.x%)

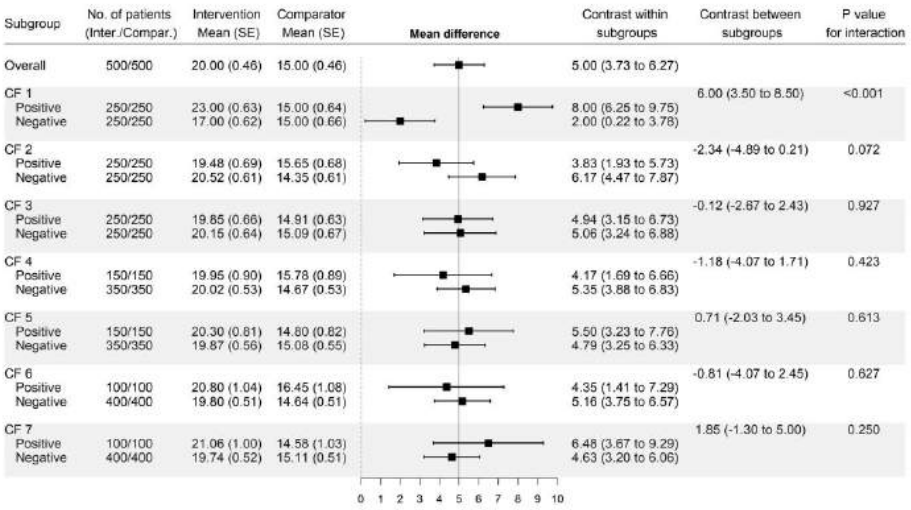
9.2 Shell figure 1. Study flow chart



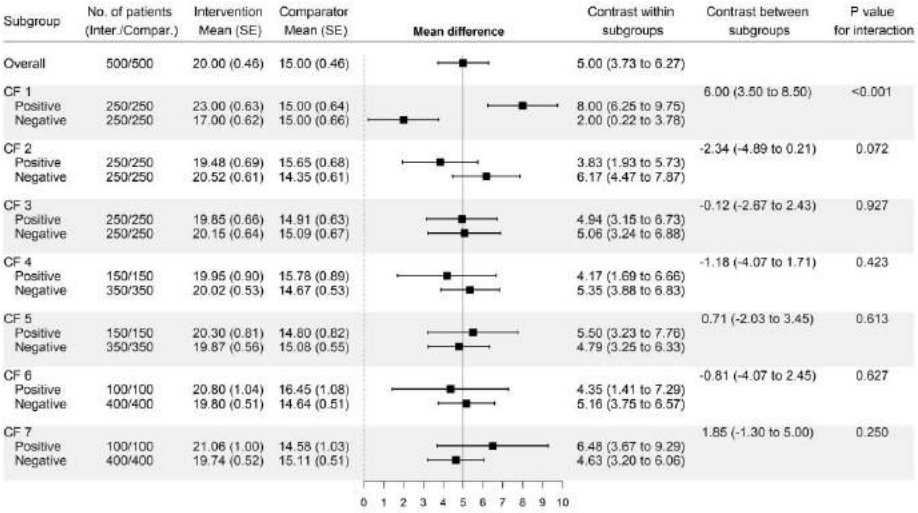
9.3 Shell figure 2. Forest plots of the Treatment effect (QE vs HE) across different subgroups at week 12 (top) and 26 (bottom) based on binary baseline variables (effect modifiers) in the ITT population.

Figure examples shown below for illustration, reprinted from Christensen et al. J Clin Epidemiol. 2021 Jun;134:174-177, distributed under the terms of the Creative Commons CC-BY license (i.e. no permissions required).

Week 12



Week 26



Abbreviations: CF, Contextual Factor (treatment effect modifier); Inter., Intervention; Compar., Comparator; SE, Standard Error

10 REFERENCES

1. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain. Am J Sports Med. 2016;44(5):1165-71.
2. Rathleff CR, Baird WN, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Hip and knee strength is not affected in 12-16 year old adolescents with patellofemoral pain--a cross-sectional population-based study. PLoS One. 2013;8(11):e79153.
3. Roush JR, Curtis Bay R. Prevalence of anterior knee pain in 18-35 year-old females. Int J Sports Phys Ther. 2012;7(4):396-401.
4. Biber R, Gregory A. Overuse injuries in youth sports: is there such a thing as too much sports? Pediatr Ann. 2010;39(5):286-92.
5. Rathleff CR, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Half of 12-15-year-olds with knee pain still have pain after one year. Dan Med J. 2013;60(11):A4725.
6. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med. 2016.
7. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'Best Practice Guide to Conservative Management of Patellofemoral Pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med. 2015;49(14):923-34.
8. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev. 2015;1:CD010387.
9. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley K, et al. Can we predict the outcome for people with patellofemoral pain? A systematic review on prognostic factors and treatment effect modifiers. Br J Sports Med. 2017;51(23):1650-60.
10. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. J Clin Epidemiol. 2021;134:174-7.
11. Hott A, Brox JI, Pripp AH, Juel NG, Liavaag S. Predictors of Pain, Function, and Change in Patellofemoral Pain. Am J Sports Med. 2020;48(2):351-8.
12. Lack S, Barton C, Vicenzino B, Morrissey D. Outcome predictors for conservative patellofemoral pain management: a systematic review and meta-analysis. Sports Med. 2014;44(12):1703-16.
13. Dolak W, Raderer M, Maresch J, Muellauer L, Puespoek A, Chott A, et al. Detection of gastric MALT lymphoma spreading to the small bowel by enteroscopy. Endoscopy. 2011;43(8):731-3.
14. Ryan D, Heaner M. Guidelines (2013) for managing overweight and obesity in adults. Preface to the full report. Obesity (Silver Spring). 2014;22 Suppl 2:S1-3.
15. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain. 2015;156(6):1003-7.
16. Steinberg N, Tenenbaum S, Zeev A, Pantanowitz M, Waddington G, Dar G, et al. Generalized joint hypermobility, scoliosis, patellofemoral pain, and physical abilities in young dancers. BMC Musculoskelet Disord. 2021;22(1):161.
17. Grahame R, Bird HA, Child A. The revised (Brighton 1998) criteria for the diagnosis of benign joint hypermobility syndrome (BJHS). J Rheumatol. 2000;27(7):1777-9.
18. Folci M, Capsoni F. Arthralgias, fatigue, paresthesias and visceral pain: can joint hypermobility solve the puzzle? A case report. BMC Musculoskelet Disord. 2016;17:58.

19. Hansen R, Lundgaard-Nielsen M, Henriksen M. Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study. PeerJ. 2021;9:e12203.
20. Herrington L. Knee valgus angle during single leg squat and landing in patellofemoral pain patients and controls. Knee. 2014;21(2):514-7.
21. Holden S, Boreham C, Doherty C, Delahunt E. Two-dimensional knee valgus displacement as a predictor of patellofemoral pain in adolescent females. Scand J Med Sci Sports. 2017;27(2):188-94.
22. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22(10):1911-20.
23. Bartholomew C, Lack S, Neal B. Altered pain processing and sensitisation is evident in adults with patellofemoral pain: a systematic review including meta-analysis and meta-regression. Scand J Pain. 2019;20(1):11-27.
24. Rathleff MS, Roos EM, Olesen JL, Rasmussen S, Arendt-Nielsen L. Lower mechanical pressure pain thresholds in female adolescents with patellofemoral pain syndrome. J Orthop Sports Phys Ther. 2013;43(6):414-21.
25. Boonstra AM, Schiphorst Preuper HR, Balk GA, Stewart RE. Cut-off points for mild, moderate, and severe pain on the visual analogue scale for pain in patients with chronic musculoskeletal pain. Pain. 2014;155(12):2545-50.
26. Maclachlan LR, Matthews M, Hodges PW, Collins NJ, Vicenzino B. The psychological features of patellofemoral pain: a cross-sectional study. Scand J Pain. 2018;18(2):261-71.
27. Domenech J, Sanchis-Alfonso V, Espejo B. Changes in catastrophizing and kinesiophobia are predictive of changes in disability and pain after treatment in patients with anterior knee pain. Knee Surg Sports Traumatol Arthrosc. 2014;22(10):2295-300.
28. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med. 2000;23(4):351-65.
29. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. J Behav Med. 1997;20(6):589-605.
30. Kortlever JT, Janssen SJ, van Berckel MM, Ring D, Vranceanu AM. What Is the Most Useful Questionnaire for Measurement of Coping Strategies in Response to Nociception? Clin Orthop Relat Res. 2015;473(11):3511-8.
31. Adams JH, Williams AC. What affects return to work for graduates of a pain management program with chronic upper limb pain? J Occup Rehabil. 2003;13(2):91-106.
32. McPoil TG, Vicenzino B, Cornwall MW, Collins N, Warren M. Reliability and normative values for the foot mobility magnitude: a composite measure of vertical and medial-lateral mobility of the midfoot. J Foot Ankle Res. 2009;2:6.
33. Tan JM, Crossley KM, Munteanu SE, Collins NJ, Hart HF, Donnar JW, et al. Associations of foot and ankle characteristics with knee symptoms and function in individuals with patellofemoral osteoarthritis. J Foot Ankle Res. 2020;13(1):57.
34. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley KM, et al. Does foot mobility affect the outcome in the management of patellofemoral pain with foot orthoses versus hip exercises? A randomised clinical trial. Br J Sports Med. 2020;54(23):1416-22.
35. Barton CJ, Bonanno D, Levinger P, Menz HB. Foot and ankle characteristics in patellofemoral pain syndrome: a case control and reliability study. J Orthop Sports Phys Ther. 2010;40(5):286-96.

Appendix 6 - Supplementary table S1-3

Table S1. Primary and Key Secondary Outcomes at week 12 and 26 in the PP population. Based on repeated measures linear mixed models, where missing data is assumed to be missing at random.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (95% CI)	P-value
Primary outcome – week 12:				
Change in AKPS questionnaire – score (0 to 100); equivalence test*	7.3 (0.9)	7.3 (0.9)	0.0 (-2.4 to 2.5)	<0.0001
Change in AKPS questionnaire – score (0 to 100); superiority test*				0.974
Key Secondary outcomes – week 12:				
Change in KOOS Pain – score (0-100)	9.5 (1.2)	6.6 (1.1)	2.9 (-0.3 to 6.1)	
Change in KOOS Function – score (0 to 100)	6.2 (1.0)	5.3 (0.9)	0.9 (-1.7 to 3.6)	
Change in KOOS Quality of life – score (0 to 100)	11.6 (1.6)	12.1 (1.5)	-0.5 (-4.7 to 3.8)	
Other Secondary outcomes – week 12:				
Change in KOOS Sports and recreation– score (0 to 100)	13.5 (1.9)	11.1 (1.8)	2.3 (-2.8 to 7.5)	
Change in KOOS Symptoms – score (0 to 100)	5.3 (0.9)	5.3 (0.9)	0.2 (-2.2 to 2.7)	
Change in isometric muscle strength				
Hip abductors (N)	14.0 (2.1)	13.4 (2.1)	0.6 (-5.1 to 6.4)	
Hip adductors (N)	11.9 (2.1)	18.0 (2.0)	-6.1 (-11.8 to -0.4)	
Hip extensors (N)	16.0 (2.8)	15.8 (2.8)	0.2 (-7.7 to 8.0)	
Hip flexors (N)	10.2 (2.2)	12.2 (2.1)	-2.1 (-8.1 to 4.0)	
Hip external rotators (N)	3.0 (5.1)	10.4 (5.0)	-7.4 (-21.4 to 6.6)	
Hip internal rotators (N)	10.1 (1.7)	13.0 (1.7)	-2.9 (-7.5 to 1.7)	
Knee extensors (quadriceps) (N)	35.9 (6.4)	35.7 (6.4)	0.2 (-17.6 to 18.0	
Knee flexors (hamstrings) (N)	45.4 (4.9)	45.7 (4.8)	-0.4 (-13.8 to 13.1)	
Change in Dynamic Assessment of Pain (VRS (0-10))	-0.7 (0.1)	-0.2 (0.1)	-0.6 (-0.9 to -0.2)	
Change in EQ5D Questionnaire (index 1,000 to -0.624)	0.066 (0.013)	0.036 (0.012)	0.031 (-0.004 to 0.065)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.0 (0.3)	2.3 (0.3)	-0.2 (-1.1 to 0.6)	
Outcomes at week 26:				
Change in AKPS questionnaire – score (0 to 100)	9.4 (0.9)	9.1 (0.9)	0.3 (-2.2 to 2.9)	
Change in KOOS Pain – score (0-100)	10.6 (1.2)	11.6 (1.2)	-1.0 (-4.4 to 2.3)	
Change in KOOS Function – score (0 to 100)	6.6 (1.0)	8.0 (1.0)	-1.4 (-4.1 to 1.4)	
Change in KOOS Quality of life – score (0 to 100)	16.6 (1.6)	20.0 (1.6)	-3.4 (-7.8 to 1.1)	
Change in KOOS Sports and recreation– score (0 to 100)	13.9 (1.9)	16.1 (1.9)	-2.3 (-7.6 to 3.0)	
Change in KOOS Symptoms – score (0 to 100)	5.8 (0.9)	7.2 (0.9)	-1.4 (-4.0 to 1.1)	
Change in EQ5D Questionnaire (index 1,000 to -0.624)	0.089 (0.013)	0.071 (0.013)	0.018 (-0.019 to 0.054)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.4 (0.3)	2.9 (0.3)	-0.5 (-1.4 to 0.4)	

*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

Table S2. Primary and Key Secondary Outcomes at week 12 and 26 in the ITT population adjusted for treatment adherence. Based on repeated measures linear mixed models, where missing data is assumed to be missing at random.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (95% CI)	P-value
Primary outcome:				
Week 12: Change in AKPS questionnaire – score (0 to 100); equivalence test*	7.5 (0.8)	7.2 (0.8)	0.3 (-1.9 to 2.5)	<0.0001
Week 12: Change in AKPS questionnaire – score (0 to 100); superiority test*				0.782
Week 26: Change in AKPS questionnaire – score (0 to 100)	9.8 (0.8)	8.9 (0.8)	0.9 (-1.4 to 3.1)	
Key Secondary outcomes:				
Week 12: Change in KOOS Pain – score (0-100)	9.4 (1.0)	6.2 (1.0)	3.1 (0.2 to 6.0)	
Week 12: Change in KOOS Function – score (0 to 100)	5.7 (0.9)	5.0 (0.9)	0.7 (-1.7 to 3.1)	
Week 12: Change in KOOS Quality of life – score (0 to 100)	10.7 (1.4)	12.1 (1.4)	-1.3 (-5.3 to 2.6)	
Other Secondary outcomes				
Week 12: Change in KOOS Sports and recreation– score (0 to 100)	13.9 (1.7)	10.8 (1.7)	3.1 (-1.6 to 7.9)	
Week 12: Change in KOOS Symptoms – score (0 to 100)	4.8 (0.8)	4.8 (0.8)	0.03 (-2.2 to 2.2)	
Week 12: Change in isometric muscle strength				
Hip abductors (N)	13.6 (1.8)	13.1 (1.9)	0.5 (-4.6 to 5.6)	
Hip adductors (N)	10.6 (1.9)	16.0 (1.9)	-5.4 (-10.6 to -0.1)	
Hip extensors (N)	16.2 (2.5)	13.4 (2.6)	2.8 (-4.2 to 9.8)	
Hip flexors (N)	11.7 (2.0)	11.1 (2.1)	0.7 (-5.0 to 6.3)	
Hip external rotators (N)	1.6 (4.3)	8.2 (4.4)	-6.6 (-18.8 to 5.5)	
Hip internal rotators (N)	9.2 (1.5)	10.3 (1.6)	-1.1 (-5.4 to 3.1)	
Knee extensors (quadriceps) (N)	32.9 (5.6)	32.7 (5.8)	0.2 (-15.6 to 16.0)	
Knee flexors (hamstrings) (N)	38.2 (4.1)	41.5 (4.1)	-3.4 (-14.7 to 7.9)	
Week 12: Change in Dynamic Assessment of Pain (VRS (0-10))	-0.8 (0.1)	-0.2 (0.1)	-0.6 (-0.9 to -0.2)	
Week 12: Change in EQ5D Questionnaire (index 1,000 to -0.624)	0.067 (0.011)	0.035 (0.011)	0.032 (0.001 to 0.063)	
Week 12: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.2 (0.3)	2.1 (0.3)	0.1 (-0.7 to 0.9)	
Week 26: Change in KOOS Pain – score (0-100)	10.7 (1.1)	10.2 (1.1)	0.5 (-2.5 to 3.5)	
Week 26: Change in KOOS Function – score (0 to 100)	6.6 (0.9)	6.9 (0.9)	-0.2 (-2.7 to 2.2)	
Week 26: Change in KOOS Quality of life – score (0 to 100)	15.9 (1.5)	18.9 (1.5)	-3.0 (-7.1 to 1.1)	
Week 26: Change in KOOS Sports and recreation– score (0 to 100)	15.2 (1.8)	14.4 (1.8)	0.8 (-4.1 to 5.7)	
Week 26: Change in KOOS Symptoms – score (0 to 100)	5.9 (0.8)	6.1 (0.8)	-0.2 (-2.5 to 2.1)	
Week 26: Change in EQ5D Questionnaire (index 1,000 to -0.624)	0.093 (0.012)	0.069 (0.012)	0.024 (-0.009 to 0.057)	
Week 26: Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.6 (0.3)	2.6 (0.3)	0.0 (-0.9 to 0.8)	

*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

Table S3. Primary and Key Secondary Outcomes at week 12 in the ITT population. Based on analysis of covariance**,** where missing data is conservatively imputed using baseline observation carried forward.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (95% CI)	P-value
Primary outcome:				
Week 12: Change in AKPS questionnaire – score (0 to 100); equivalence test*	6.5 (0.8)	6.4 (0.8)	0.1 (-2.3 to 2.4)	<0.0001
Week 12: Change in AKPS questionnaire – score (0 to 100); superiority test*				0.952
Key Secondary outcome:				
Week 12: Change in KOOS Pain – score (0-100)	8.2 (1.1)	5.8 (1.1)	2.5 (-0.7 to 5.7)	
Week 12: Change in KOOS Function – score (0 to 100)	5.0 (0.9)	4.6 (0.9)	0.4 (-2.2 to 2.9)	
Week 12: Change in KOOS Quality of life – score (0 to 100)	9.5 (1.5)	11.0 (1.5)	-1.4 (-5.5 to 2.7)	

*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.

Appendix 7 – Adverse events

Table S4. Adverse events in the intention-to-treat population

	QE (n=100)	HE (n=100)
Exposure time – patient weeks	900	948
AE - no. of patients (%)	15 (15%)	16 (16%)
AE - no. of events (rate – event per patient week)	21 (0.03)	19 (0.02)
AEs leading to discontinuation - no. of patients (%)	1 (1%)	1 (1%)
Maximum reported severity of AEs, no. of patients (%)		
Mild	9 (9%)	8 (8%)
Moderate	4 (4%)	6 (6%)
Severe	2 (2%)	2 (2%)
AEs, relationship to trial treatment, no. of events (rate – event per patient week)		
Not related	2 (0.002)	2 (0.002)
Probably not related	3 (0.003)	6 (0.006)
Probably related	16 (0.02)	11 (0.01)
AEs, classification, no. of events (rate – event per patient week)		
PFP pain exacerbation	8 (0.01)	6 (0.01)
Muscle soreness	12 (0.01)	12 (0.01)
Other	1 (0.001)	1 (0.001)
Deaths - no. of events (rate – event per patient week)	0 (0)	0 (0)
AE; Adverse event. The severity of an adverse event refers to the maximum intensity of the event. An event was considered severe (compared with mild or moderate) if it interfered substantially with the patient's usual activities.		

Appendix 8 – Cross tabulated agreements (Study 3)

Supplementary file for Study 3: Agreement of visual assessment of dynamic knee joint alignment in patients with patellofemoral pain.

Table 1. Intrarater agreement Forward Lunge - raw scores

		rater 1									
		-4	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**Total**
rater 1	**-4**	1	2	0	0	0	0	0	0	0	3
	-3	0	6	2	0	0	0	0	0	0	8
	-2	0	3	4	4	1	0	0	0	0	12
	-1	0	1	5	10	2	0	0	0	0	18
	0	0	0	1	3	15	0	0	0	0	19
	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0
	Total	1	12	12	17	18	0	0	0	0	60

Table 2. Intrarater agreement Forward Lunge – categorical classification

		rater 1			
		Varus	**Neutral**	**Valgus**	**Total**
rater 1	**Varus**	0	0	0	0
	Neutral	0	30	7	37
	Valgus	0	5	18	23
	Total	0	35	25	60

Table 3. Interrater agreement Forward Lunge - raw scores

		rater 2									
		-4	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**Total**
rater 1	**-4**	0	3	0	0	0	0	0	0	0	3
	-3	0	4	3	1	0	0	0	0	0	8
	-2	0	2	4	6	0	0	0	0	0	12
	-1	0	0	5	10	3	0	0	0	0	18
	0	0	1	2	5	10	1	0	0	0	19
	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0
	Total	0	10	14	22	13	1	0	0	0	60

Table 4. Interrater agreement Forward Lunge – categorical classification

		rater 2			
		Varus	**Neutral**	**Valgus**	**Total**
rater 1	**Varus**	0	0	0	0
	Neutral	0	29	8	37
	Valgus	0	7	16	23
	Total	0	36	24	60

Table 5. Intrarater agreement single leg squat - raw scores

		rater 1									
		-4	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**Total**
rater 1	**-4**	5	1	0	0	0	0	0	0	0	6
	-3	3	11	4	1	0	0	0	0	0	19
	-2	0	4	8	4	0	0	0	0	0	16
	-1	0	0	2	6	2	0	0	0	0	10
	0	0	0	1	4	3	1	0	0	0	9
	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0
	Total	8	16	15	15	5	1	0	0	0	60

Table 6. Intrarater agreement single leg squat - categorical classification

		rater 1			
		Varus	**Neutral**	**Valgus**	**Total**
rater 1	**Varus**	0	0	0	0
	Neutral	0	16	3	19
	Valgus	0	5	36	41
	Total	0	21	39	60

Table 7. Interrater agreement single leg squat - raw scores

		rater 2									
		-4	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**Total**
rater 1	**-4**	1	3	1	0	1	0	0	0	0	6
	-3	2	9	4	4	0	0	0	0	0	19
	-2	0	5	4	7	0	0	0	0	0	16
	-1	0	1	4	4	1	0	0	0	0	10
	0	0	0	4	2	2	1	0	0	0	9
	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0
	Total	3	18	17	17	4	1	0	0	0	60

Table 8. Interrater agreement single leg squat - categorical classification

		rater 2			
		Varus	**Neutral**	**Valgus**	**Total**
rater 1	**Varus**	0	0	0	0
	Neutral	0	10	9	19
	Valgus	0	12	29	41
	Total	0	22	38	60

Study 1 – Quadriceps or hip exercises for patellofemoral pain? A randomized controlled equivalence trial

Notes:

In the paper we refer to the study protocol in the supplemental material. This corresponds to the study protocol found in Appendix 1.

In the 'Interventions' section we refer to the complete description of the interventions and the progression principles. These correspond to the exercise description in Appendix 2. Further, we refer to the information leaflet which corresponds to the leaflet in Appendix 3.

In the 'Statistical analysis' section we refer to the statistical analysis plan in the supplements. This corresponds to the statistical analysis plan provided in Appendix 4.

In the 'Key secondary and other secondary outcomes' section we refer to the sensitivity analyses, which correspond to the supplementary tables S1-3 in Appendix 6.

Quadriceps or hip exercises for patellofemoral pain? A randomized controlled equivalence trial

Rudi Hansen[1], Christoffer Brushøj[2,] Michael Skovdal Rathleff[3,4], Stig Peter Magnusson[2,5], Marius Henriksen[6]

1. Department of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
2. Institute of Sports Medicine Copenhagen, Department of Orthopedic Surgery, Copenhagen University Hospital – Bispebjerg and Frederiksberg, Copenhagen, Denmark
3. Department of Health Science and Technology, Aalborg University, Aalborg, Denmark.
4. Center for General Practice at Aalborg University, Aalborg, Denmark.
5. Center for Healthy Aging, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
6. The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

Corresponding author: Rudi Hansen, Department of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark.
Email: rudi.hansen@regionh.dk
Word count: 3778

ABSTRACT

Objective: To assess effectiveness equivalence between two commonly prescribed 12-week exercise programs targeting either the quadriceps or the hip muscles in patients with patellofemoral pain (PFP).

Methods: This randomized controlled equivalence trial included 200 participants with a clinical diagnosis of PFP. Participants were randomly assigned to either a 12-week quadriceps-focused (QE) or a hip-focused (HE) exercise program. The primary outcome was the change in Anterior Knee Pain Scale (AKPS) (0-100) from baseline to 12-week follow-up. Prespecified equivalence margins of ±8 points on the AKPS were chosen to demonstrate comparable effectiveness. Key secondary outcomes were the Knee Injury and Osteoarthritis Outcome Score questionnaire (KOOS) pain, physical function, and knee-related quality of life subscales.

Results: The least squares mean changes in AKPS (primary outcome) were 7.5 for QE and 7.2 for HE (difference 0.3 points, 95% CI −1.9 to 2.4; test for equivalence $p<0.0001$). None of the group differences in key secondary outcomes exceeded predefined equivalence margins.

Conclusion: 12-week focused quadriceps and hip focused exercise protocols were equivalent in changes in symptoms and function for patients with PFP.

INTRODUCTION

Patellofemoral pain (PFP) is a common knee problem, with point prevalences from 6 to 7 % in adolescents and up to 13% in young adults (1-3). More than one in two with PFP report persistent pain after 5–8 years (4), with an associated frequent use of pain killers, a lower physical activity level, and low quality of life (1, 4-6). Recent systematic reviews and a network meta-analysis recommend exercise therapy (mainly comprising exercises for the hip, the knee, or both the hip and knee) for improving pain and function in people with PFP (7-10). However, these studies also underline the uncertainty about which type of exercises that are most effective for PFP. Despite the latest consensus document on managing PFP recommends including hip exercises, direct comparisons of exercise protocols are few (9), with short intervention follow-up periods (11), and with sample sizes insufficient to detect clinically relevant differences in outcomes (12, 13). Collectively, this challenges the choice of the most appropriate treatment and may also explain the variation in clinical practice (14). Hence, there is a need for large high-quality studies of comparative effectiveness of quadriceps and hip muscle exercises for PFP. Accordingly, the aim of this study is to assess effectiveness equivalence between a focused “Quadriceps Exercise” (QE) protocol and a focused “Hip Exercise” (HE) protocol on symptoms and function in patients with PFP.

METHODS

Study design

In this single-center randomized, controlled, assessor-blinded, equivalence trial with two parallel intervention groups we compared a QE and HE protocol. Evaluations and assessments took place at the Department of Physical and Occupational Therapy at Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark at baseline and 12 weeks. Further, the participants were invited to an online collection of patient-reported outcomes 26 weeks after baseline. The trial design is illustrated in the supplements together with the trial protocol. Ethical approval was obtained from the Health Research Ethics Committee of the Capital region, Denmark (H-16045755). This report follows the CONSORT extension for non-pharmacological treatments guideline and the TIDieR checklist for intervention description (15, 16). The study was registered prospectively at www.ClinicalTrials.gov on March 3, 2017 (NCT03069547).

Patient and Public Involvement statement

Patients were not engaged in the development stages of the study nor in the conduct or oversight of the study. All participants were offered a layman resume of results and conclusion of the study by email.

Participants

Between April 10, 2017 and December 3, 2021 participants were recruited from the Institute of Sports Medicine Copenhagen (ISMC), Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark. Inclusion was halted for 8 weeks from March 12, 2020 due to the COVID-19 pandemic. ISMC is a medical unit for patients with injuries in the musculoskeletal system caused by participation in sports activities. Patients are referred to ISMC from primary care physicians. All participants underwent a clinical examination by a specialist in sports medicine but were not screened for eligibility using radiographs or other imaging. All participants provided written informed consent before participation.

Inclusion criteria were a clinical diagnosis of PFP in at least one knee confirmed by an experienced sports medicine physician, average knee pain during activities of daily living in the last week of ≥3/10 on a verbal rating scale, insidious onset of symptoms unrelated to trauma, persistent pain for at least four weeks, and anterior knee pain associated with at least three of the following: During or after activity, prolonged sitting, stair ascent or descent, or squatting. The main exclusion criteria were other knee conditions, including meniscal or other intra-articular injuries to the knee, history of recurrent patellar subluxation or dislocation, and previous knee surgery (complete details are provided in the study protocol in the supplements). Potential participants were informed about the trial during an interview with a sports medicine physician, and after at least 24 hours of consideration an investigator obtained written informed consent and coordinated trial visits. The most symptomatic knee at baseline was chosen as the study knee.

Randomization and blinding

Before randomization, demographic information and all baseline measures were obtained. Participants were randomly assigned (1:1) in permuted blocks of 4 and 6 (randomly distributed) according to a computer-generated list of random numbers, to one of the two groups (QE or HE). Individual allocations were concealed in sealed opaque envelopes, stored in a locked cupboard without access for investigators or outcome assessor, and delivered sequentially to the study

physiotherapist at randomization. The clinical staff delivering the interventions and participants were not blinded to treatment allocation. The investigators and the outcome assessor were blinded to allocation, and participants were requested not to disclose allocation during clinical assessments.

Interventions

Both the hip and knee focused exercise programs were inspired by previous research (17) and followed recommended prescribing guidelines (18, 19). The exercise interventions lasted for 12 weeks with three weekly home-based exercise sessions consisting of three sets of 8-12 repetitions. Key parameters of the exercise programs are shown in table 1, and the complete description of the interventions is provided in the supplementary file.

The hip exercise (HE) program consisted of hip external rotation (clam shell), side-lying/standing hip abduction, and prone/standing hip extension. The HE exercises were chosen due to their documented activation of the hip abductors, external rotators, and hip extensors (20-23), wide use in clinical practice, and because they do not strain the patellofemoral joint excessively. *The quadriceps exercise (QE) program* consisted of sitting knee extension, squat, and forward lunge. The exercises has been shown effective in recruiting the quadriceps muscle (22) and appear effective in the treatment of PFP (24, 25).

Both exercise programs were initiated at an individual clinical visit. An experienced physiotherapist introduced the participant to the allocated exercise program (QE or HE) and provided instructions to the individual exercises. Elastic bands, free weights, and body weight were used to provide resistance. The participants were informed to perform 8-12 repetitions in each set. The last repetitions should be difficult to perform while still allowing the participant to maintain high quality of movement (i.e., full range of motion and without any compensatory movements (judged by the physiotherapist)) throughout the entire program. The participants were instructed to increase resistance whenever they could complete 14 repetitions in a set. This was emphasized during the instructional session and during each follow-up visit (19) (progression principles are specified in the supplementary file). The exercise programs included monthly clinical supervision visits. Reduction in the exercise load (ROM, number of sets/repetitions) could be made in case of significant knee pain exacerbations.

All participants – irrespective of group allocation – received the information leaflet "Managing my patellofemoral pain" containing general information on possible causes and management of PFP. The leaflet is available in the supplementary file. Further a comprehensive

exercise leaflet with guidance on the exercises, progression/regression, and pain management was handed out. All the physiotherapists involved in the study (n=5) were instructed to communicate in the same way, and training sessions were held in the planning stage to ensure standardization of communication and practice.

Adherence to the prescribed exercise protocol was monitored by a self-administered exercise diary, which the participants were encouraged to fill in after each exercise session. The participants were asked to record date, number of repetitions and sets for each exercise, and the resistance (i.e., elastic band color corresponding to a specified resistance or weights in kg). The exercise diary was brought at the monthly clinical visits to optimize compliance and handed in at the 12-weeks assessment. The criteria for satisfactory intervention adherence in both groups was 24 of the 36 scheduled training sessions (66%).

Primary outcome

The primary outcome was change from baseline in the Anterior Knee Pain Scale (AKPS) questionnaire (26) at week 12. The AKPS questionnaire is a widely used and well-validated questionnaire for assessing the severity of symptoms and physical limitations in people with PFP (26). The 13 items in the questionnaire are summed up to give a total score ranging from 0 to 100, with high scores indicating less symptoms. The minimal clinically important change is established at 8-10 points (27).

Key secondary and other secondary outcomes

Key secondary outcomes were changes from baseline in the Knee Injury and Osteoarthritis Outcome Score questionnaire (KOOS) pain, physical function, and knee-related quality of life subscales (28). Other secondary outcomes included changes from baseline in the KOOS sports/recreation and symptoms subscales, Pain Self-Efficacy Questionnaire (29), the EuroQoL EQ-5D-3L Questionnaire (30), assessment of pain on a 0–10 numeric rating scale (NRS) during activity (30 s of performing repeated deep knee-bends from a standing position) (31), and global perceived effect on overall health, pain and function measured on a 15-point Likert scale ranging from -7 (much worse) to +7 (much better). Further, isometric muscle strength of hip abductors, hip adductors, hip external rotators, hip internal rotators, hip extensors, hip flexors, knee flexors, and knee extensors were measured with a handheld dynamometer. The testing was performed in a clinical examination room with the participant lying or sitting on an examination table with and

without external fixation according to published and validated protocols (32-34). Three consecutive isometric maximal contractions were performed with a 30-s rest period between each trial, and the maximum value was used for analysis. Changes from baseline in the patient reported outcomes (questionnaires) at week 26 were also recorded.

Sample size

The sample size was calculated to allow for test of equivalence of the treatment groups at 90% power and an alpha level of 0.05 using a two one-sided test (one-sided alpha of 0.025) with equivalence margins of ±8 AKPS points, assuming a mean difference of 0 points and a common standard deviation of 15 points (35, 36). From this, 77 patients were required in each treatment group. To account for a dropout rate of approximately 20% the sample size was a priori increased to 100 participants in each group.

Statistical analysis

The analysis was performed according to the a priori statistical analysis plan that was published at www.clinicaltrials.gov on January 24, 2022 (before last data recording from the last participant; see the supplements).

The primary analysis was performed using the intention-to-treat population; patients were assessed and analyzed as members of their randomized groups, irrespective of adherence to the planned course of treatment. Continuous outcomes were analyzed as change from baseline using repeated measures linear mixed models. Participants were included as random effects (normal distribution assumed), and group (2 levels), week (including all timepoints (3 levels; week 0, 12 and 26)), and the corresponding interaction was included as fixed effects. An adjustment for baseline values was done to increase precision. Assumptions underlying the linear mixed models (e.g., normality of residuals) were assessed. Results are reported as least squares means and standard errors (SE), and differences between least squares means are reported with two-sided 95% confidence intervals (CI). The group difference in the primary outcome was assessed for equivalence by a two one-sided test of equivalence with alpha 0.025 assessing if the 95% CI respects the predefined equivalence margin of ±8 AKPS points corresponding to the established cut-off value for making the distinction between improved or unimproved (27). No explicit adjustments for multiplicity were applied, rather the key secondary outcome measures were analyzed in a prioritized order. Missing values for items in the AKPS and Pain Self-Efficacy

Questionnaire were substituted with the arithmetic mean of values from the available items. If more than 25% of items were missing, the outcome was regarded missing for the patient (13, 37). For the KOOS questionnaire, a mean score for each subscale was calculated, as long as at least 50% of the items were answered for each subscale. If more than 50% of the subscale items were omitted, the response was considered invalid. Imputation of missing item values for the EuroQoL EQ-5D-3L Questionnaire, was handled according to the user guide (euroqol.org). Imputation of missing values in AKPS constituted less than 5% of all questionnaire data. Complete missing data were handled implicitly in the intention-to-treat analysis by the linear mixed models (38). Sensitivity analyses (39) were performed for the primary and key secondary outcomes at week 12 by repeating the primary analyses on the per-protocol population predefined as participants with satisfactory adherence and without major protocol deviations. Further, we repeated the primary analysis (on the ITT population without imputation of missing data) adjusting for treatment adherence (percentage). Finally, we performed an analysis of covariance of the primary and key secondary outcomes at week 12 (i.e., without the repeated measures) on the ITT population with a conservative non-responder baseline observation carried forward imputation of missing data at week 12 adjusted for the baseline values. For all the secondary sensitivity analyses assumptions underlying the models (e.g., normality of residuals) were assessed. If the primary analysis and the sensitivity analyses confirm each other, confidence in the results is increased both regarding equivalence and superiority claims. All analyses were performed in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Participants

From April 10th 2017 through December 3rd 2021, 288 individuals were screened for eligibility (figure 1); 88 were ineligible for inclusion. Thus, 200 participants underwent randomization; 100 were assigned to QE and 100 to HE. The mean age was 27.2 years (SD 6.4); 69% were females; and the mean BMI was 22.6 (SD 3.0). Baseline characteristics were similar in the two groups (table 2). Participants completed on average 28 (77%) training sessions out of 36 possible sessions. During the course of the intervention, 6 participants (4 QE; 2 HE) had alterations to their allocated exercise programs (mainly reduced ROM in the weight bearing exercises) and 8 (5 QE; 3 HE) had number of sets and repetitions reduced due to knee pain exacerbations.

Primary outcome

The mean changes in AKPS questionnaire score from baseline to week 12 were 7.5 (CI 6.0 to 9.0) in the QE group and 7.2 (CI 5.7 to 8.7) in the HE group (group difference: 0.3 points, 95% CI −1.9 to 2.4; p=0.804 for test of superiority). The 95% CI of the group difference in change in AKPS questionnaire from baseline to week 12 was within the predefined equivalence margin of±8 points (p<0.0001 for equivalence, table 3). The trajectories of the AKPS questionnaire are shown in figure 2.

Key secondary and other secondary outcomes

For the key secondary outcomes, the estimated treatment differences between groups at week 12 were 3.0 points (95% CI 0.1 to 5.9) for KOOS pain score, 0.6 points (95% CI −1.7 to 3.0) for KOOS function, and -1.5 points (95% CI −5.4 to 2.5) for KOOS quality of life score. The key secondary outcomes all respected the predefined criteria for equivalence, although the between group difference for KOOS pain was statistically significant in favor of QE (table 3). Finally, the results in the primary and key secondary outcomes appeared unchanged at week 26 (table 5). There were no statistically or clinically significant differences between groups in the other secondary, safety and exploratory outcomes at week 12 (table 3) and week 26 (table 5). The overall pattern of results for all outcomes was unchanged in the sensitivity analyses (online supplemental tables S1-S3).

Safety

Adverse events were typically mild to moderate, mostly related to muscle soreness, and were similar in the two groups (table 4). Severe adverse events that gave interference with the participants' usual activities were exacerbation of knee pain (n=2), headache (n=1), and back pain (n=1).

DISCUSSION

The latest systematic review with network meta-analysis demonstrated uncertainty about which exercises to include when managing patients with PFP (7). The results of this study provide much needed evidence to inform clinical practice and highlight that an exercise program that focused on either quadriceps or hip muscles provided equivalent improvements in symptoms and function in

the short (12 weeks) and medium term (26 weeks). Treatment adherence was similar in the two groups as were adverse events that were few.

Our results support recently published RCTs comparing hip and knee focused exercise protocols. In Hott et al (13), 112 patients were randomized to three groups (a 6-week intervention consisting of patient education combined with isolated hip-focused exercise, traditional knee-focused exercise, or free physical activity); the data indicated no difference in the primary outcome AKPS between groups. This is in line with previous studies, showing no difference in pain and function at 6-8 weeks between a hip and a knee exercise group (40, 41). Three studies have found hip exercises to be more effective than knee-focused exercise (42-44); however, the sample sizes were typically quite modest (15-18 per group), and one study lacked randomization. This study is the first using an equivalence design that allows us to draw reliable conclusions regarding the comparative effectiveness of hip and knee focused exercises for PFP. Hence, our results extend current understanding and effectively demonstrate equivalent effectiveness of hip and knee focused exercise for PFP.

Both exercise programs were associated with improvements in AKPS score (7.5 points for QE and 7.2 points for HE), but the improvement did not reach the established minimally clinical important change threshold. The within-group changes for QE and HE are similar to those previously reported (13), but are somewhat lower than those reported in other RCTs evaluating the effect of hip and knee exercises in adolescents and adults with PFP (35, 40, 41, 45-47). This difference may be explained by the setting of this study. Patients included in this study were referred to specialized rehabilitation most often due to long-standing symptoms, which is reflected in the patient demographics. Previous studies have shown that long symptom duration is associated with worse outcomes (irrespective of treatment) (4, 48, 49) which may explain the somewhat small within group changes. Mean pain duration in this study was higher when compared to most studies that report on this (40, 45, 47). Another plausible explanation for the small within group changes could be differences in attention and supervision during the intervention period compared to other studies. Most of the interventions in comparable studies were supervised, but this is not always feasible in a clinical setting. The patient–physiotherapist relationship and the overall healthcare setting are relevant categories of contextual factors that may modify treatment effects (50).

Both groups had 10-11% improvements in hip abduction and knee extension muscle strength after the 12-week training period irrespective of group allocation, which is similar to previous studies (13, 40, 44). Since some aspects of the hip exercises involve weight bearing,

several other muscles are recruited when performing the exercise, including the quadriceps. Likewise for the quadriceps focused exercises, a possible parallel training of the hip (and other synergistic) muscles cannot be completely ruled out. One could argue that this may explain the lack of group difference in the outcomes. However, in a large randomized clinical trial with 218 participants with PFP, increases in muscle strength did not mediate improvements in pain (51). This suggests that improvements in muscle strength might not be the driver of beneficial outcomes, and that other mechanisms are more important.

The somewhat modest improvements seen in our and other recent studies on exercises for PFP raise the question if treatment plans focusing on strengthening and biomechanically informed movement quality alone address the right components contributing to the pain experience. Growing evidence suggests that psychological features may play a role in long-standing PFP (52-54). Future studies should aim at identifying possible patient characteristics that predict successful outcomes.

Clinical Implications

We found that quadriceps exercises and hip exercises are equally effective treatments in the management of patients with PFP. The is in line with the most recent consensus document that recommends combining quadriceps and hip exercise (9). This implies that therapists should use their clinical reasoning and include patient preferences when designing an exercise rehabilitation program for the individual patient. Such shared decision may improve healthcare efficiency and is recommended in the rehabilitation of patients with PFP (9, 55, 56). However, although personalization of exercise interventions to individual patients or subgroup of patients may be a useful strategy that can ultimately lead to improved outcomes for patients (57), such strategy remains to be supported by research evidence – preferably from prospective randomized trials.

Limitations and strengths

There are inherent limitations to this study. First, the exercise programs were home-based with limited supervision, which may introduce a risk that the exercises were not performed correctly. While more regular visits to the clinician would assure adherence and fidelity to the treatments, this would not be in accordance our intention to resemble a clinical setting, where multiple weekly visits are not feasible (14). Second, the exercise adherence data was based on self-reporting, which introduces an inherent risks of overestimation due to social desirability, recall period, and selective recall (58). Third, this study was a single center trial which may limit the

external validity. Fourth, as the main part of the interventions were unsupervised, contamination (deliberate switch of exercise program) may have occurred. Finally, as part of a prognostic sub-study (to be published separately) the physiotherapists recorded the participants' projected prognosis after the first clinical encounter but not disclosed to the participants. As this was done post randomization it may have introduced some expectation biases with the physiotherapists. However, such prognoses are inherent in clinical interactions between patients and health care providers and thus this does not represent deviations from normal clinical practice. The strengths of this trial included the relatively large sample size and the equivalence design, which increase the precision of the estimated group differences, and the reporting of adverse events. Furthermore, this is the first study comparing hip and knee exercises with an intervention period of 12 weeks, with comparable studies ranging from 3 to 8 weeks of intervention (10, 11).

CONCLUSION

In individuals with patellofemoral pain, 12-week quadriceps-focused and hip-focused exercise programs provided equivalent effectiveness for improvements in symptoms and function.

SUMMARY BOX

What is already known on this topic

Current evidence supports exercise therapy in the treatment of patients with patellofemoral pain. However, there is uncertainty about the comparative effectiveness of hip and knee exercises and high quality is needed to guide clinical practice.

What this study adds

This study adds to the current knowledge-base and demonstrates that quadriceps focused exercises and hip focused exercises provide equivalent benefits for patients with patellofemoral pain.

How this study might affect research, practice, or policy

Based on this study, clinicians can include patient preferences and individualisation in the choice of either hip or knee focused exercises in the management of patients with patellofemoral pain.

Acknowledgements

We thank the participants who volunteered to the study, and the physiotherapists at the Department of Physical and Occupational Therapy at Bispebjerg-Frederiksberg Hospital who delivered treatments**.**

Contributors

RH: concept/design, data collection, data analysis, manuscript draft, critical revision. MH: concept/design, data analysis, critical revision, guarantor. CB: concept/design, medically responsible, eligibility screening, critical revision. MSR: concept/design, critical revision. SPM: critical revision.

Funding

There was no external funding.

Competing interests

None declared.

Patient consent for publication

Not relevant

Ethics approval

This study involves human participants and was approved by the Committees on Health Research Ethics, Capital Region (H-16045755 (approved December 15, 2016)). Participants gave informed consent to participate in the study before taking part.

Data availability statement

Data are available upon reasonable request.

FIGURES AND TABLES

Figure 1. CONSORT flow diagram

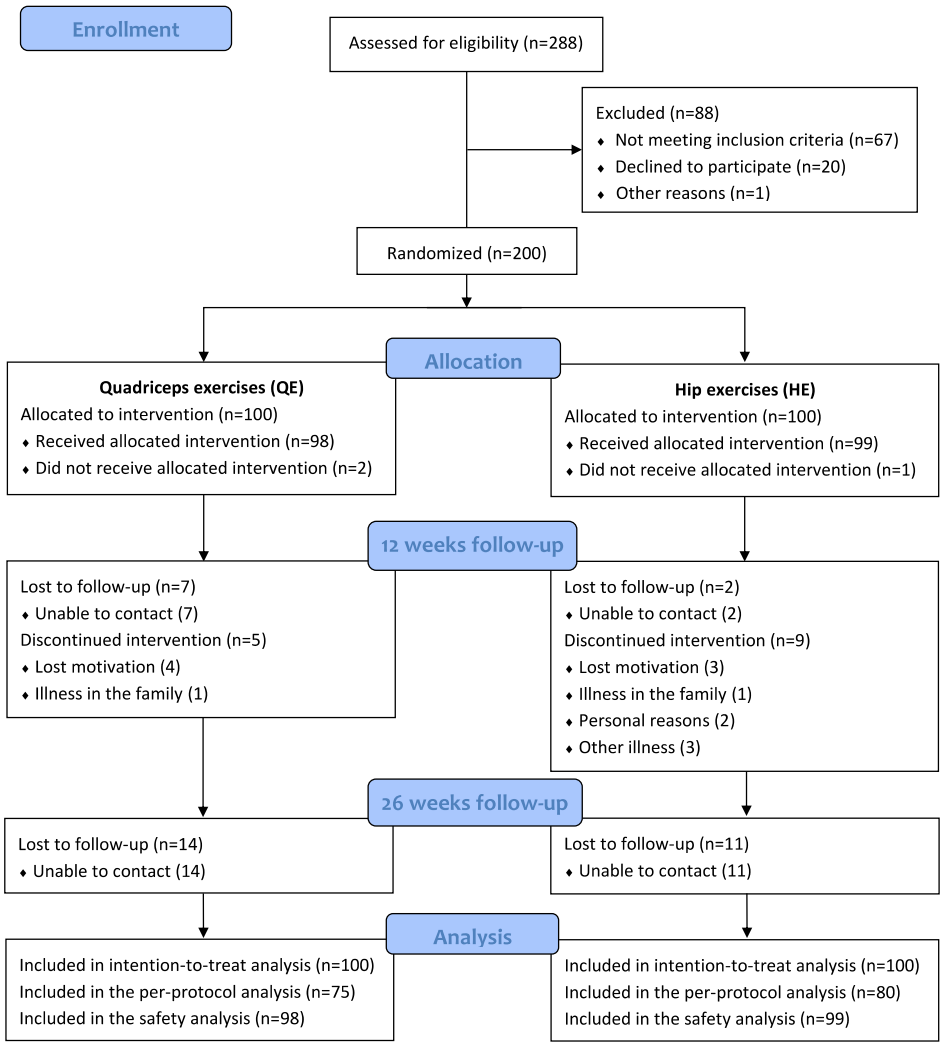


Figure 2. Trajectories of the AKPS questionnaire in the ITT population. High values represent high levels of self-reported function; low values represent low levels of self-reported function. Data points represent least squares means; error bars. SE

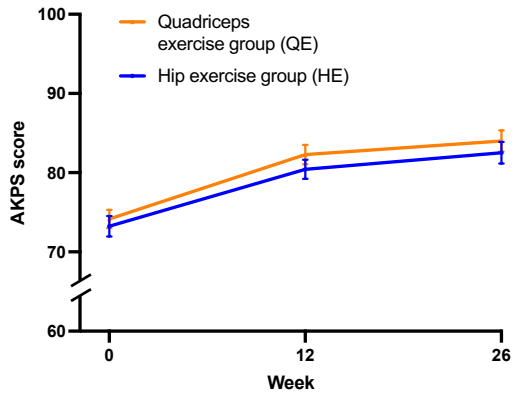


Table 1. Key parameters of the exercise programs

	Number of repetitions/sets	Time under tension	Rest in between sets	Means of progression	Number of exercise interventions per week
QE-1: Sitting leg extension	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands on ankles	3/week
QE-2: Squat	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding weight in a backpack (e.g. sand, flour, bottles of water) or by holding dumbbells in the hands.	3/week
QE-3: Lunge	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	As above	3/week
HE-1: Clam-shell	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands just above both knees	3/week
HE-2: Side-lying/standing hip abduction	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	As above	3/week
HE-3: Standing hip extension	8-12 reps./3 sets	2-4 sec/repetition	1 min. 30 sec – 2 min.	Adding elastic bands from underneath the foot to the knee of the moving limb	3/week

QE: Quadriceps exercise
HE: Hip exercise
Both groups were instructed to warm up by performing 20 repetitions of exercise QE-1 (for QE) or HE-1 (for HE) without external load.
Both groups were instructed to increase resistance whenever the participants were able to perform 2 repetitions more than the desired number (i.e., 14 or more)

Table 2. Demographics and Baseline Characteristics

	Quadriceps exercise group (QE) N=100	**Hip exercise group (HE) N=100**
	Mean (SD)	**Mean (SD)**
Demographics		
Age, years	27.2 (6.3)	27.2 (6.7)
Female sex (n[%])	66 (66%)	72 (72%)
Body mass, kg	68.2 (12.4)	67.6 (13.0)
Height, cm	172.4 (8.5)	173.2 (10.7)
Body Mass Index, BMI (kg/m2)	22.8 (3.01)	22.4 (2.9)
Symptom duration, months	47.3 (49.4)	52.8 (54.1)
Symptoms duration, months (median (IQR))*	36 (48)	30 (60)
AKPS questionnaire score (0-100)	74.2 (11.6)	73.3 (13.0)
KOOS (0-100)		
Pain	70.8 (15.6)	72.2 (14.1)
Physical Function	84.1 (13.2)	83.4 (13.1)
Symptoms	79.6 (14.0)	80.7 (13.3)
QoL	44.4 (15.1)	44.2 (14.7)
Sports & Recreation	56.7 (24.9)	59.3 (24.4)
Dynamic assessment of pain (VRS 0-10)	1.9 (2.2)	1.8 (1.8)
Dynamic assessment of pain (VRS 0-10) (median (IQR))*	1.0 (3.5)	2.0 (3.0)
Isometric muscle strength		
Hip abductors (N)	129.5 (40.9)	129.5 (41.2)
Hip adductors (N)	121.2 (40.6)	122.4 (47.3)
Hip extensors (N)	175.3 (46.8)	181.0 (56.0)
Hip flexors (N)	189.1 (55.8)	194.3 (63.0)
Hip external rotators (N)	101.1 (30.2)	100.7 (37.7)
Hip internal rotators (N)	123.9 (100.9)	109.4 (42.8)
Knee extensors (quadriceps) (N)	299.2 (113.1)	292.0 (121.9)
Knee flexors (hamstrings) (N)	316.6 (117.1)	302.8 (129.5)
Pain Self-efficacy questionnaire (0-60)	47.5 (8.6)	46.8 (9.8)
EuroQoL EQ5D Questionnaire (-0.624 to 1.000)	0.755 (0.175)	0.757 (0.127)

Values are presented as means and standard deviations (SD) unless otherwise stated.
* Both means (SD) and medians (IQR) are presented as data is not normally distributed.
IQR: Inter quartile range
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

Table 3. Primary and Secondary Outcomes at week 12 in the ITT population. Based on repeated measures linear mixed models, where missing data is assumed to be missing at random.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (95% CI)	P-value
Primary outcome:				
Change in AKPS questionnaire score (0 to 100); equivalence test*	7.5 (0.8)	7.2 (0.8)	0.3 (-1.9 to 2.4)	<0.0001
Change in AKPS questionnaire score (0 to 100); superiority test*				0.804
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	9.4 (1.0)	6.4 (1.0)	3.0 (0.1 to 5.9)	
Change in KOOS Function – score (0 to 100)	5.7 (0.9)	5.1 (0.9)	0.6 (-1.7 to 3.0)	
Change in KOOS Quality of life – score (0 to 100)	10.7 (1.4)	12.2 (1.4)	-1.5 (-5.4 to 2.5)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	13.8 (1.7)	11.0 (1.7)	2.8 (-1.9 to 7.6)	
Change in KOOS Symptoms – score (0 to 100)	4.8 (0.8)	4.9 (0.8)	-0.1 (-2.3 to 2.1)	
Change in isometric muscle strength				
Hip abductors (N)	13.7 (1.8)	13.3 (1.9)	0.4 (-4.7 to 5.5)	
Hip adductors (N)	10.7 (1.9)	16.3 (1.9)	-5.5 (-10.8 to -0.3)	
Hip extensors (N)	16.4 (2.5)	13.7 (2.5)	2.6 (-4.4 to 9.6)	
Hip flexors (N)	11.8 (2.0)	11.2 (2.0)	0.6 (-5.0 to 6.2)	
Hip external rotators (N)	1.7 (4.3)	8.4 (4.4)	-6.7 (-18.8 to 5.4)	
Hip internal rotators (N)	9.4 (1.5)	10.6 (1.5)	-1.3 (-5.5 to 3.0)	
Knee extensors (quadriceps) (N)	33.3 (5.6)	33.3 (5.7)	-0.1 (-15.8 to 15.7)	
Knee flexors (hamstrings) (N)	37.6 (4.1)	42.1 (4.1)	-4.4 (-15.8 to 6.9)	
Change in Dynamic Assessment of Pain (VRS (0-10))	-0.8 (0.1)	-0.2 (0.1)	-0.6 (-0.9 to -0.3)	
Change in EQ5D Questionnaire (index -0.624 to 1.000)	0.067 (0.011)	0.035 (0.011)	0.024 (-0.009 to 0.057)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.1 (0.3)	2.1 (0.3)	0.0 (-0.8 to 0.8)	
Treatment adherence				
Treatment adherence (%)	75.0 (23.2)	79.0 (21.3)	-4.0 (-10.2 to 2.2)	
Treatment adherers (adherence ≥66%) - no. (%)	82 (82.0%)	85 (85.0%)		

*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
AKPS: Anterior Knee Pain Scale
KOOS: Knee injury and osteoarthritis outcome score.
VRS: Verbal Rating Scale

Table 4. Adverse events in the intention-to-treat population

	QE (n=100)	HE (n=100)
Exposure time – patient weeks	900	948
AE - no. of patients (%)	15 (15%)	16 (16%)
AE - no. of events (rate – event per patient week)	21 (0.03)	19 (0.02)
AEs leading to discontinuation - no. of patients (%)	1 (1%)	1 (1%)
Maximum reported severity of AEs, no. of patients (%)		
Mild	9 (9%)	8 (8%)
Moderate	4 (4%)	6 (6%)
Severe	2 (2%)	2 (2%)
AEs, relationship to trial treatment, no. of events (rate – event per patient week)		
Not related	2 (0.002)	2 (0.002)
Probably not related	3 (0.003)	6 (0.006)
Probably related	16 (0.02)	11 (0.01)
AEs, classification, no. of events (rate – event per patient week)		
PFP pain exacerbation	8 (0.01)	6 (0.01)
Muscle soreness	12 (0.01)	12 (0.01)
Other	1 (0.001)	1 (0.001)
Deaths - no. of events (rate – event per patient week)	0 (0)	0 (0)
AE; Adverse event. The severity of an adverse event refers to the maximum intensity of the event. An event was considered severe (compared with mild or moderate) if it interfered substantially with the patient's usual activities.		

Table 5. Primary and Secondary Outcomes at week 26 in the ITT population. Based on repeated measures linear mixed models, where missing data is assumed to be missing at random.

	QE (N=100) Mean (SE)	HE (N=100) Mean (SE)	Mean difference (CI)	P-value
Primary outcome:				
Change in AKPS questionnaire – score (0 to 100); equivalence test*	9.8 (0.8)	9.0 (0.8)	0.9 (-1.4 to 3.1)	<0.0001
Change in AKPS questionnaire – score (0 to 100); superiority test*				0.449
Key Secondary outcome:				
Change in KOOS Pain – score (0-100)	10.7 (1.1)	10.4 (1.1)	0.4 (-2.7 to 3.4)	
Change in KOOS Function – score (0 to 100)	6.6 (0.9)	6.9 (0.9)	-0.3 (-2.8 to 2.1)	
Change in KOOS Quality of life – score (0 to 100)	15.9 (1.5)	19.0 (1.5)	-3.2 (-7.2 to 0.9)	
Other Secondary Outcomes:				
Change in KOOS Sports and recreation– score (0 to 100)	15.1 (1.8)	14.6 (1.8)	0.5 (-4.4 to 5.4)	
Change in KOOS Symptoms – score (0 to 100)	5.9 (0.8)	6.2 (0.8)	-0.3 (-2.6 to 2.0)	
Change in EQ5D Questionnaire (index -0.624 to 1.000)	0.093 (0.012)	0.069 (0.012)	0.024 (-0.009 to 0.057)	
Transition Questionnaire of global perceived change in overall health, pain, and function (Likert scale -7 to 7)	2.6 (0.3)	2.7 (0.3)	-0.1 (-0.9 to 0.7)	

Group values for QE and HE are presented as least squares means ± standard error.
Mean differences are presented as least squares means and 95% confidence intervals (CI).
*Primary outcome was analyzed using both a test for equivalence and a test for superiority.
KOOS: Knee injury and osteoarthritis outcome score.

REFERENCES

1. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain. Am J Sports Med. 2016;44(5):1165-71.
2. Roush JR, Curtis Bay R. Prevalence of anterior knee pain in 18-35 year-old females. Int J Sports Phys Ther. 2012;7(4):396-401.
3. Smith BE, Selfe J, Thacker D, Hendrick P, Bateman M, Moffatt F, et al. Incidence and prevalence of patellofemoral pain: A systematic review and meta-analysis. PLoS One. 2018;13(1):e0190892.
4. Lankhorst NE, van Middelkoop M, Crossley KM, Bierma-Zeinstra SM, Oei EH, Vicenzino B, et al. Factors that predict a poor outcome 5-8 years after the diagnosis of patellofemoral pain: a multicentre observational analysis. Br J Sports Med. 2016;50(14):881-6.
5. Rathleff MS, Winiarski L, Krommes K, Graven-Nielsen T, Holmich P, Olesen JL, et al. Pain, Sports Participation, and Physical Function in Adolescents With Patellofemoral Pain and Osgood-Schlatter Disease: A Matched Cross-sectional Study. J Orthop Sports Phys Ther. 2020;50(3):149-57.
6. Andreucci A, Roos EM, Rasmussen S, Olesen JL, Holmich P, Thorborg K, et al. Analgesic use in adolescents with patellofemoral pain or Osgood-Schlatter Disease: a secondary cross-sectional analysis of 323 subjects. Scand J Pain. 2021.
7. Winters M, Holden S, Lura CB, Welton NJ, Caldwell DM, Vicenzino BT, et al. Comparative effectiveness of treatments for patellofemoral pain: a living systematic review with network meta-analysis. Br J Sports Med. 2020.
8. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'Best Practice Guide to Conservative Management of Patellofemoral Pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med. 2015;49(14):923-34.
9. Collins NJ, Barton CJ, van Middelkoop M, Callaghan MJ, Rathleff MS, Vicenzino BT, et al. 2018 Consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th International Patellofemoral Pain Research Retreat, Gold Coast, Australia, 2017. Br J Sports Med. 2018;52(18):1170-8.
10. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev. 2015;1:CD010387.
11. Na Y, Han C, Shi Y, Zhu Y, Ren Y, Liu W. Is Isolated Hip Strengthening or Traditional Knee-Based Strengthening More Effective in Patients With Patellofemoral Pain Syndrome? A Systematic Review With Meta-analysis. Orthop J Sports Med. 2021;9(7):23259671211017503.
12. Saltychev M, Dutton RA, Laimi K, Beaupre GS, Virolainen P, Fredericson M. Effectiveness of conservative treatment for patellofemoral pain syndrome: A systematic review and meta-analysis. J Rehabil Med. 2018;50(5):393-401.
13. Hott A, Brox JI, Pripp AH, Juel NG, Paulsen G, Liavaag S. Effectiveness of Isolated Hip Exercise, Knee Exercise, or Free Physical Activity for Patellofemoral Pain: A Randomized Controlled Trial. Am J Sports Med. 2019;47(6):1312-22.
14. Smith BE, Hendrick P, Bateman M, Moffatt F, Rathleff MS, Selfe J, et al. Current management strategies for patellofemoral pain: an online survey of 99 practising UK physiotherapists. BMC Musculoskelet Disord. 2017;18(1):181.

15. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.
16. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg. 2012;10(1):28-55.
17. Holden S, Rathleff MS, Jensen MB, Barton CJ. How can we implement exercise therapy for patellofemoral pain if we don't know what was prescribed? A systematic review. Br J Sports Med. 2018;52(6):385.
18. Toigo M, Boutellier U. New fundamental resistance exercise determinants of molecular and cellular muscle adaptations. Eur J Appl Physiol. 2006;97(6):643-63.
19. American College of Sports M. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.
20. Boren K, Conrey C, Le Coguic J, Paprocki L, Voight M, Robinson TK. Electromyographic analysis of gluteus medius and gluteus maximus during rehabilitation exercises. Int J Sports Phys Ther. 2011;6(3):206-23.
21. Distefano LJ, Blackburn JT, Marshall SW, Padua DA. Gluteal muscle activation during common therapeutic exercises. J Orthop Sports Phys Ther. 2009;39(7):532-40.
22. Ekstrom RA, Donatelli RA, Carp KC. Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises. J Orthop Sports Phys Ther. 2007;37(12):754-62.
23. MacAskill MJ, Durant TJ, Wallace DA. Gluteal muscle activity during weightbearing and non-weightbearing exercise. Int J Sports Phys Ther. 2014;9(7):907-14.
24. Kooiker L, Van De Port IG, Weir A, Moen MH. Effects of physical therapist-guided quadriceps-strengthening exercises for the treatment of patellofemoral pain syndrome: a systematic review. J Orthop Sports Phys Ther. 2014;44(6):391-402, B1.
25. Herrington L, Al-Sherhi A. A controlled trial of weight-bearing versus non-weight-bearing exercises for patellofemoral pain. J Orthop Sports Phys Ther. 2007;37(4):155-60.
26. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy. 1993;9(2):159-63.
27. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? Arch Phys Med Rehabil. 2004;85(5):815-22.
28. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther. 1998;28(2):88-96.
29. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain. 2007;11(2):153-63.
30. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33(5):337-43.
31. Klokker L, Christensen R, Waehrens EE, Bandak E, Bartholdy C, Bliddal H, et al. Dynamic weight-bearing assessment of pain in knee osteoarthritis: construct validity, responsiveness, and interpretability in a research setting. Health Qual Life Outcomes. 2016;14:91.
32. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. Scand J Med Sci Sports. 2010;20(3):493-501.

33. Romero-Franco N, Jimenez-Reyes P, Montano-Munuera JA. Validity and reliability of a low-cost digital dynamometer for measuring isometric strength of lower limb. J Sports Sci. 2017;35(22):2179-84.
34. Mentiplay BF, Perraton LG, Bower KJ, Adair B, Pua YH, Williams GP, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. PLoS One. 2015;10(10):e0140822.
35. Fukuda TY, Rossetto FM, Magalhaes E, Bryk FF, Lucareli PR, de Almeida Aparecida Carvalho N. Short-term effects of hip abductors and lateral rotators strengthening in females with patellofemoral pain syndrome: a randomized controlled clinical trial. J Orthop Sports Phys Ther. 2010;40(11):736-42.
36. van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. BMJ. 2009;339:b4074.
37. Chavance M. Handling Missing Items in Quality of Life Studies. Communications in Statistics - Theory and Methods. 2004;33(6):1371-83.
38. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. JAMA. 2016;315(4):407-8.
39. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012;367(14):1355-60.
40. Ferber R, Bolgla L, Earl-Boehm JE, Emery C, Hamstra-Wright K. Strengthening of the hip and core versus knee muscles for the treatment of patellofemoral pain: a multicenter randomized controlled trial. J Athl Train. 2015;50(4):366-77.
41. Saad MC, Vasconcelos RA, Mancinelli LVO, Munno MSB, Liporaci RF, Grossi DB. Is hip strengthening the best treatment option for females with patellofemoral pain? A randomized controlled trial of three different types of exercises. Braz J Phys Ther. 2018;22(5):408-16.
42. Dolak KL, Silkman C, Medina McKeon J, Hosey RG, Lattermann C, Uhl TL. Hip strengthening prior to functional exercises reduces pain sooner than quadriceps strengthening in females with patellofemoral pain syndrome: a randomized clinical trial. J Orthop Sports Phys Ther. 2011;41(8):560-70.
43. Khayambashi K, Fallah A, Movahedi A, Bagwell J, Powers C. Posterolateral hip muscle strengthening versus quadriceps strengthening for patellofemoral pain: a comparative control trial. Arch Phys Med Rehabil. 2014;95(5):900-7.
44. Baldon Rde M, Serrao FV, Scattone Silva R, Piva SR. Effects of functional stabilization training on pain, function, and lower extremity biomechanics in women with patellofemoral pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2014;44(4):240-51, A1-A8.
45. Fukuda TY, Melo WP, Zaffalon BM, Rossetto FM, Magalhaes E, Bryk FF, et al. Hip posterolateral musculature strengthening in sedentary women with patellofemoral pain syndrome: a randomized controlled clinical trial with 1-year follow-up. J Orthop Sports Phys Ther. 2012;42(10):823-30.
46. Almeida GPL, Rodrigues H, Coelho BAL, Rodrigues CAS, Lima POP. Anteromedial versus posterolateral hip musculature strengthening with dose-controlled in women with patellofemoral pain: A randomized controlled trial. Phys Ther Sport. 2021;49:149-56.
47. Constantinou A, Mamais I, Papathanasiou G, Lamnisos D, Stasinopoulos D. Comparing hip and knee focused exercises versus hip and knee focused exercises with the use of blood flow restriction training in adults with patellofemoral pain. Eur J Phys Rehabil Med. 2022;58(2):225-35.

48. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley K, et al. Can we predict the outcome for people with patellofemoral pain? A systematic review on prognostic factors and treatment effect modifiers. Br J Sports Med. 2017;51(23):1650-60.
49. Hott A, Brox JI, Pripp AH, Juel NG, Liavaag S. Predictors of Pain, Function, and Change in Patellofemoral Pain. Am J Sports Med. 2020;48(2):351-8.
50. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet. 2001;357(9258):757-62.
51. Holden S, Matthews M, Rathleff MS, Kasza J, Fohx G, Vicenzino B. How Do Hip Exercises Improve Pain in Individuals With Patellofemoral Pain? Secondary Mediation Analysis of Strength and Psychological Factors as Mechanisms. J Orthop Sports Phys Ther. 2021;51(12):602-10.
52. Maclachlan LR, Collins NJ, Matthews MLG, Hodges PW, Vicenzino B. The psychological features of patellofemoral pain: a systematic review. Br J Sports Med. 2017;51(9):732-42.
53. Sigmund KJ, Bement MKH, Earl-Boehm JE. Exploring the Pain in Patellofemoral Pain: A Systematic Review and Meta-Analysis Examining Signs of Central Sensitization. J Athl Train. 2021;56(8):887-901.
54. Vicenzino BT, Rathleff MS, Holden S, Maclachlan L, Smith BE, de Oliveira Silva D, et al. Developing Clinical and Research Priorities for Pain and Psychological Features in People With Patellofemoral Pain: An International Consensus Process With Health Care Professionals. J Orthop Sports Phys Ther. 2022;52(1):29-39.
55. Barton CJ, Crossley KM. Sharing decision-making between patient and clinician: the next step in evidence-based practice for patellofemoral pain? Br J Sports Med. 2016;50(14):833-4.
56. Legare F, Stacey D, Turcotte S, Cossi MJ, Kryworuchko J, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev. 2014(9):CD006732.
57. Selfe J, Janssen J, Callaghan M, Witvrouw E, Sutton C, Richards J, et al. Are there three main subgroups within the patellofemoral pain population? A detailed characterisation study of 127 patients to help develop targeted intervention (TIPPs). Br J Sports Med. 2016;50(14):873-80.
58. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-7.

Study 2 – Differential effects of quadriceps and hip muscle exercise programs for patellofemoral pain: A secondary effect modifier analysis of a randomized trial

Notes:

In the 'Participants' section we refer to the full description of in- and exclusion criteria in the supplemental material. This corresponds to the study protocol found in Appendix 1.

In the 'Interventions' section we refer to the complete description of the interventions. This corresponds to the exercise description in Appendix 2. Further, we refer to the information leaflet which corresponds to the leaflet in Appendix 3.

In the section 'Outcome and potential effect modifiers' we refer to the full description of the characteristics. These are found in the statistical analysis plan of Study 2 in Appendix 5.

In 'Statistical analyses' we also refer to the statistical analysis plan provided in Appendix 5.

Differential effects of quadriceps and hip muscle exercise programs for patellofemoral pain: A secondary effect modifier analysis of a randomized trial

Rudi Hansen[1,2], Christoffer Brushøj[2,] Michael Skovdal Rathleff[3,4], Stig Peter Magnusson[1,2], Marius Henriksen[5]

1. Department of Physical and Occupational Therapy, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
2. Institute of Sports Medicine Copenhagen, Department of Orthopedic Surgery, Copenhagen University Hospital – Bispebjerg and Frederiksberg, Copenhagen, Denmark and Center for Healthy Aging, Department of Clinical Medicine, University of Copenhagen, Denmark
3. Department of Health Science and Technology, Aalborg University, Aalborg, Denmark.
4. Center for General Practice at Aalborg University, Aalborg, Denmark.
5. The Parker Institute, Copenhagen University Hospital – Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

ABSTRACT

Objective: To identify baseline characteristics that modify the effect of 12-week quadriceps vs. hip muscle exercise programs in patients with patellofemoral pain (PFP).

Methods: A secondary analysis of a 26-week randomized trial including 200 participants with PFP. Participants were randomly assigned to a 12-week quadriceps-focused (QE) or hip-focused (HE) exercise program. Outcome was change from baseline in the Anterior Knee Pain Scale (AKPS) at week 12 and 26. Subgroups were predefined and based on baseline information: Presence of low back, hip, ankle, or bilateral knee pain, body mass index (BMI), sex, age, education, occupation, hypermobility, quadriceps strength, dynamic knee alignment, midfoot mobility, exercise self-efficacy, pain self-efficacy, pain catastrophizing, neuropathic pain, pain duration, and pain severity.

Results: Participants with signs of pain catastrophizing seemed to benefit more from QE with a subgroup difference in treatment effect of 8.3 AKPS points at week 12 (95%CI 1.6 to 15.0; P=0.016). At week 26, participants with a baseline BMI above 25 seemed to benefit more from QE with a subgroup difference in treatment effect of 11.1 (95%CI 4.8 to 17.8; P=0.001), and participants with severe knee pain at baseline seem to benefit from the HE protocol with a subgroup difference of -9.1 (95%CI -15.7 to -2.6; P=0.006). The identified subgroups were small and the results imprecise.

Conclusion: QE may potentially provide more benefits than HE among PFP patients with signs of pain catastrophizing or overweight, whereas HE may provide more benefits than QE for patients with more severe knee pain. The results are more indicative than conclusive and need to be confirmed.

INTRODUCTION

Patellofemoral Pain (PFP) is a common knee problem, which particularly affects adolescents and young adults (1-4). PFP is characterized by pain around or behind the patella that significantly impacts function and quality of daily life (5). The majority of patients still report pain after 2 to 8 years with extensive burdens on the individual and society (6). Resistance training for the quadriceps and the hip muscles has repeatedly been shown beneficial for pain and physical function and is unequivocally recommended as a core component of the management of PFP (7-9). In general, different types of exercises produce similar small to moderate beneficial effects in pain and physical function (10-12). However, the PFP population is very heterogeneous and "one-size-fits-all"-approaches are presumably sub-optimal. This heterogeneity probably explains the overall limited beneficial effects of exercise, and the lack of differences in direct comparisons of different exercise types. Our recent randomised controlled trial (RCT) comparing quadriceps-focused and hip-muscle focused exercise programs (the COMPETE trial) showed equivalent and small to moderate improvements in symptoms and physical limitations in both groups without between-group differences (13), emphasising the need for identification of subgroups that respond differentially to these two commonly used exercise regimens.

Identifying subgroups of individuals who may benefit more from one treatment than the other, and potential treatment effect modifiers is an important goal in health research (14, 15). Extensive effort has been made to identify indicators of prediction for conservative PFP management (6, 16, 17). Several factors have been linked to a poor outcome, with longer duration of symptoms, older age, lower function, bilateral symptoms, greater difference in side-to-side knee extension strength, and number of pain sites at baseline being the most reported (16, 18-22). However, although prognostic factors help predict the likelihood of an outcome within a certain time period, they cannot predict the likelihood of an outcome after a specific treatment (16). Adequately powered RCTs with evaluation of effect modifiers as a primary aim are clearly warranted to provide clinicians with robust evidence and facilitate evidence-informed, tailored intervention to this heterogeneous patient population (14, 17, 23, 24).

The aim of this secondary pre-specified analysis was to identify contextual factors that modify the observed treatment effect of treatment (Quadriceps Exercise program (QE) vs Hip muscle Exercise program (HE)) across patient subgroups, i.e. whether the treatment effect is modified by the value of a variable assessed at baseline.

METHODS

Study design

This is a secondary analysis of a single center, randomized trial with two parallel intervention groups comparing treatment efficacy of a quadriceps exercise (QE) protocol and a hip exercise (HE) protocol with a primary endpoint at 12 weeks (after treatment) and a follow-up at 26 weeks. A prespecified aim was to explore candidate patient characteristics that predict differential responses to the two exercise programs on self-reported pain and physical function in individuals with PFP. A total of 200 patients was randomly assigned on a 1:1 basis to one of the two treatments, QE or HE. Baseline characteristics were registered and assessed as part of the baseline assessment (before randomization) in the main trial. The trial was prospectively registered in the ClinicalTrials.gov database on March 3, 2017 (NCT03069547) and approved by the Health Research Ethics Committee for the Capital Region of Denmark (H-16045755).

Participants

Between April 10, 2017, and December 3, 2021, participants were recruited from the Institute of Sports Medicine Copenhagen (ISMC), Bispebjerg-Frederiksberg Hospital, Copenhagen, Denmark. All participants provided written informed consent before participation.

Inclusion criteria were a clinical diagnosis of PFP, average knee pain during activities of daily living in the previous week of ≥3/10 on a verbal rating scale, insidious onset of symptoms unrelated to trauma and persistent for at least 4 weeks, and anterior knee pain associated with at least 3 of the following: During or after activity, prolonged sitting, stair ascent or descent, or squatting. Main exclusion criteria were other knee conditions, including meniscal or other intra-articular injuries to the knee (complete details are provided in the supplements). Potential participants were informed about the trial by a sports medicine doctor, and subsequently an investigator obtained written informed consent and coordinated trial inclusion. The most symptomatic knee at baseline was chosen as the study knee.

Randomization and blinding

Before randomization, demographic information and all baseline characteristics were obtained. Participants were randomly assigned (1:1) in permuted blocks of 4 and 6 according to a computer-

generated list of random numbers, to one of the two groups (QE or HE). The allocation was concealed in opaque envelopes, stored in a locked cupboard without access for investigators or outcome assessor, and delivered sequentially to the study physiotherapist at randomization. The clinical staff delivering the interventions and participants were not blinded to treatment allocation. The investigators and the outcome assessor were blinded to allocation where possible, and participants were requested not to disclose allocation during clinical assessments.

Interventions

The exercise interventions were inspired by previous research (25) and adhered to recommended prescribing guidelines (26, 27). The intervention period was 12 weeks with three weekly home-based exercise sessions consisting of three sets of 8-12 repetitions. *The quadriceps exercise (QE) program* consisted of sitting knee extension, squat, and forward lunge. *The hip exercise (HE) program* consisted of hip external rotation (clam shell), side-lying/standing hip abduction, and prone/standing hip extension. Both exercise programs were initiated at an individual clinical visit. An experienced physiotherapist introduced the participant to the allocated exercise program (HE or QE) and provided instructions to the individual exercises. Elastic bands, free weights, and body weight were used as resistance. The exercise programs included monthly clinical supervision visits. A complete description of the interventions is provided in the supplementary file.

All participants – irrespective of group allocation – received the information leaflet "Managing my patellofemoral pain" containing general information on causes and management of PFP. The leaflet is available in the supplementary file.

Outcome and potential effect modifiers

The primary outcome was change from baseline in the Anterior Knee Pain Scale (AKPS) questionnaire at week 12. The AKPS questionnaire is a widely used and well-validated questionnaire for assessing the severity of symptoms and physical limitations in people with PFP (28). The 13 items in the questionnaire are summed up to give a total score ranging from 0 to 100, with high scores indicating less symptoms. The minimal clinically important change is established at 8 points (29).

A priori we identified a range of baseline variables to be explored as potential effect modifiers. The candidate baseline characteristics encompass self-reported information as well as

clinical observations and tests and were chosen based on findings in previous studies and clinical experience (16, 17). The characteristics were dichotomized in accordance with established clinically relevant cut-off values where applicable. For certain items, i.e., quadriceps strength and age, no meaningful cut-off values were found, and dichotomization was hence based on median values. The baseline characteristics are presented in table 1. A full description of the characteristics including rationales for selecting the specific items and cut-off values is available in the supplementary files.

Statistical analysis

The statistical analyses were prespecified in a statistical analysis plan that was written and closed before data analyses (supplement). The primary analysis was performed using the intention-to-treat population; patients were assessed and analyzed as members of their randomized groups, irrespective of adherence to the planned course of treatment. The analyses focused on the interaction between presence/absence of the patient characteristics and group allocation (QE vs. HE) at each time point (week 12 and 26). Changes from baseline in AKPS were analyzed using repeated measures mixed linear models with group (QE vs HE), week (12 and 26), and candidate baseline predictor characteristic (present/absent) as fixed factors, and participant as random factor. The models were adjusted for the outcome baseline value. Missing data were handled implicitly by the mixed linear models, and the models were adjusted for the baseline AKPS value. All analyses were performed using the statistical software SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Between April 10th 2017 and December 3rd 2021, 288 individuals were screened for eligibility (figure 1); 88 were ineligible for inclusion. Thus, 200 subjects underwent randomization; 100 were assigned to QE and 100 to HE. The mean age was 27.2 years; 66% were females; and the mean BMI was 22.6. Baseline characteristics and the distribution of the subgroups were similar in the two groups (table 2). Participants completed on average 28 (77%) training sessions, out of possible 36 (range 0–36) sessions.

The overall mean changes in AKPS score from baseline to week 12 were 7.2 points (SE 0.8) in the QE group and 7.5 points (SE 0.8) in the HE group with a group difference of -0.3 points (95% confidence interval, -2.4 to 1.9; P=0.80), which is identical to the main trial results (13). At

the 26-week follow-up, the group mean changes were 9.0 (SE 0.8) and 9.8 (SE 0.8) in the QE and HE groups, respectively, with a mean difference of -0.9 (95% CI -3.1 to 1.4; P=0.45)

There was a statistically significant subgroup difference at 12 weeks in favour of QE among participants with the baseline characteristic "Pain catastrophizing" (n=22) compared to those without signs of pain catastrophising at baseline (subgroup difference 8.3 AKPS points (95%CI 1.6 to 15.0; P=0.016). At 26 weeks, participants with a baseline BMI ≥25 m/kg^2 (n=32) seemed to benefit from QE compared to those with BMI <25 m/kg^2 (n=168) with a subgroup difference of 11.1 AKPS points (95%CI 4.8 to 17.4; P=0.001). In contrast, participants with severe baseline knee pain (n=28) seemed to benefit from HE when compared to those with mild-moderate baseline knee pain (n=172) with a subgroup difference of -9.1 AKPS points (95%CI -15.7 to -2.6; P=0.006). The results of the subgroup analyses are shown in figure 2 for the week 12 and figure 3 for the 26-week follow-up assessment.

DISCUSSION

In this secondary effect modifier analysis of an RCT, we aimed to identify baseline characteristics that could modify the effect of quadriceps focused exercises compared to exercises focusing on the hip muscles on AKPS in patients with PFP. Our results indicate that in the short term (immediately after intervention; 12 weeks) patients with self-reported signs of pain catastrophizing at baseline seem to benefit more from the QE program. Further, in the medium-term follow-up (26 weeks from baseline) participants with BMI above 25 kg/m^2 also seem to benefit from QE, whereas those with severe baseline knee pain seem to benefit from HE.

The mechanisms by which presence of signs of pain catastrophizing at baseline seem to modify the 12-week treatment effect of QE vs HE are not possible to assess in the present study. Pain catastrophizing is defined as a maladaptive cognitive-affective response to pain that involves negative thinking regarding the pain experience (30). It is a prevalent psychologic feature in patients with patellofemoral pain (17-19) and a larger psychosocial impact is associated with worse pain and reduced physical function (31, 32). Notably, the treatment modification was depleted after 26 weeks, suggesting only short-term effect modification immediately after the intervention.

The mechanisms underlying 26-week treatment effects (BMI≥25 benefit from QE and severe baseline knee pain benefit from HE) are also not possible to identify from the present data. Although high BMI has been linked to the development of PFP in adults and is a well-recognized risk factor for incidence and progression of knee OA (33, 34), no significant link has been

established between BMI and specific intervention outcomes (35, 36). A high BMI has been linked to the development of PFP in adults and is a well-recognized risk factor for knee OA (33, 34). But no association has been established between BMI and specific intervention outcomes (35). An post-hoc analysis of a randomized controlled trial investigating the treatment effect of knee exercises compared to open-label placebo in individuals with knee OA did not find a high BMI to modify the effect (37). Intuitively, a high body mass would increase patellofemoral joint loads and stress during weight bearing exercises like squat and lunges (QE), contradictive of what present results suggest.

Severe pain (above 6 on the NRS scale) was associated with beneficial effects on symptoms and function in the HE group after 26 weeks. Pain severity has been identified as a prognostic factor in PFP populations but not as an effect modifier for specific treatments (16, 18). Severe cases often describe aggravated pain with activities that include weight-bearing knee flexion, such as running, stair climbing, squatting, and jumping (23). In other musculoskeletal disorders, exercising painful joints has been found to aggravate pain (induce hyperalgesia) (38, 39), whereas exercising a distant non-painful joint have been associated with exercise-induced hypoalgesia (40, 41). This could suggest that in patients with severe pain clinicians should focus on hip exercises early in the rehabilitation, similar to what has been proposed in the field of chronic pain (41).

The absence of robust effect modifiers for PFP is in accordance with other studies comparing specific treatments. In a recently conducted RCT, patients with greater midfoot width mobility did not have superior benefits using foot orthoses, compared to hip exercises at 12 weeks follow-up (42). Further, in a secondary analysis of an RCT comparing the effectiveness of supervised exercise therapy to usual care for 6 weeks in patients with PFP, none of the tested variables had a significant interaction with treatment (24). Two factors, however, tended to have a predictive value in favor of exercise therapy: duration of complaints and sex. Patellofemoral pain is a heterogeneous condition; persons with PFP do not all have the same impairments, and not all persons with PFP respond to the same interventions (8, 43). This is likely an inherent reason for the absence of identified characteristics in current research evidence (16).

The exact mechanisms of the effect of exercise therapy on PFP are not well understood and the choice of the potential effect modifiers was therefore based on indications from the literature and clinical experience. It is likely that other factors may be at play. We did not, for example, include activity level or sports participation. However, an interaction between sport intensity and treatment was not found in an earlier study (24). Other potential effect modifiers also include pre-

clinical data, e.g., imaging, blood samples, genetic testing etc. We chose baseline characteristics that were clinically feasible and required only simple clinical assessments or patient-reported data.

The overall effect and the within group improvements in symptoms and function were modest, likely because of the long symptom duration (average above 47 months) in the included sample (13). This could challenge the detection of significant effect modifiers. Furthermore, the treatments provided in the study were both active interventions, and they resemble each other in terms of intensity, exercise type, and information given, likely also encumbering identification of subgroups with differential responses.

Clinical implications

This study does not provide robust guidance for evidence-based choice of either quadriceps or hip muscles focused exercises in the management of PFP. Together with the main findings from the trial (equivalence between QE and HE), it could be speculated that a combination program could be optimal, which aligns with current recommendations (9). The program could then be more or less focused on quadriceps or hips according to the potential effect modifiers found in the present study (acknowledging the uncertainty of the findings) and patient preferences. Shared decision may improve healthcare efficiency and is recommended in the rehabilitation of patients with PFP (9, 44, 45). The identification of the subgroups (signs of pain catastrophizing, overweight, and pain severity) is clinically feasible as they rely on simple patient reported outcomes and measurements of height and weight.

Implications for future research

The lack of effect modifiers to guide treatment in the clinic is apparent (16), and the search should continue. One solution to this problem may be the use of individual patient data (IPD) from multiple trials to perform a meta-analysis. However, differences in exercise programs, population samples, and outcome measures as well as insufficiently reported study details could raise a problem in analyzing IPD, and future studies should therefore strive to conform with research strategies in consensus statements (9, 23) and with reporting guidelines (46). A less explored avenue is to investigate preclinical assessments, such as imaging, biochemistry, genetic testing etc., but as mentioned such assessments are not feasible in the clinical settings.

Limitations and strengths

There are limitations to this study. When a set of potential factors are tested simultaneously within the same study, the overall type I error rate is increased, potentially resulting in an increased risk of a false-positive finding. As this is an explorative study, we did not correct for multiplicity or small subgroups, rather we emphasize that results should be interpretated with caution. Moreover, the identified subgroups that seem to have differential treatment responses are characterized by being small and with subgroup differences being imprecisely estimated (wide 95% CIs), and the analyses may not have been sufficiently powered to detect or reject a significant interaction. This renders the results very imprecise and inconclusive. Hence, the identified potential treatment effect modifiers are merely indications that should be explored further.

Also, dichotomizing continuous data assumes that the effect occurs only at a specific threshold, with a subsequent loss of information and statistical power. However, grouping of data helps the clinical interpretation and transferability. Most of the subgroups were defined based on thresholds established in the literature, while some characteristics were split arbitrarily, i.e., by medians, as no clinically meaningful division were feasible. Finally, it could be speculated that the interaction between continuous baseline characteristics and the effect is non-linear, which could be explored in future studies.

An apparent strength is the relatively large sample population included in the RCT and the prespecified aim to assess patient characteristics associated with a successful outcome, that allowed for the most comprehensive analysis of effect modifiers of exercise in PFP to date. Also, the prespecified statistical analysis plan strengthens the study.

CONCLUSION

We identified three patient characteristics that potentially modified the outcome of a quadriceps-focused exercise program compared to a hip-focused program in patients with PFP. Patients with signs of pain catastrophizing and a BMI ≥25, a quadriceps focused exercise program may potentially provide more benefits than a hip muscle focused exercise program, whereas a hip focused program may potentially be better for patients with more severe knee pain at baseline. Because of small subgroups and risk of multiplicity, the results are more indicative than conclusive and need to be confirmed.

Figure 1. CONSORT flow diagram

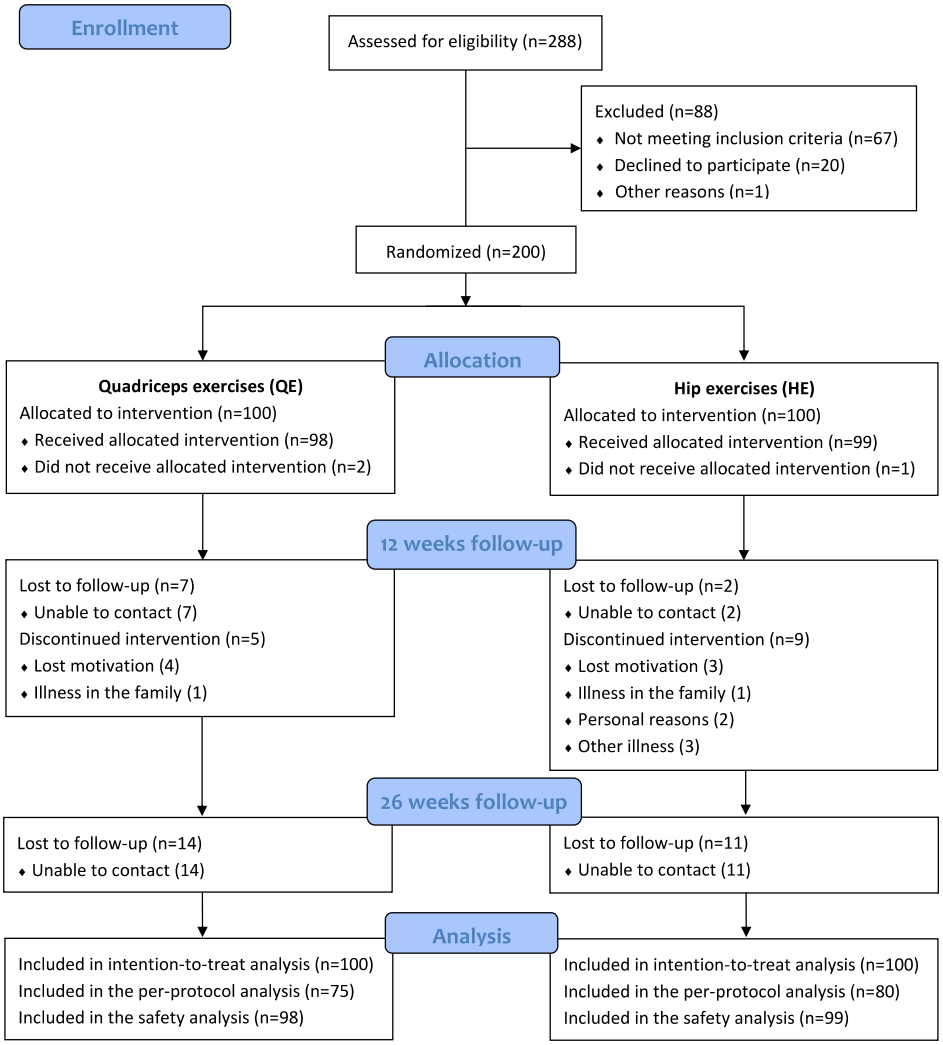


Figure 2. Forest plots of the Treatment effect (QE vs HE) across different subgroups at week 12 based on dichotomous baseline variables (effect modifiers) in the ITT population.

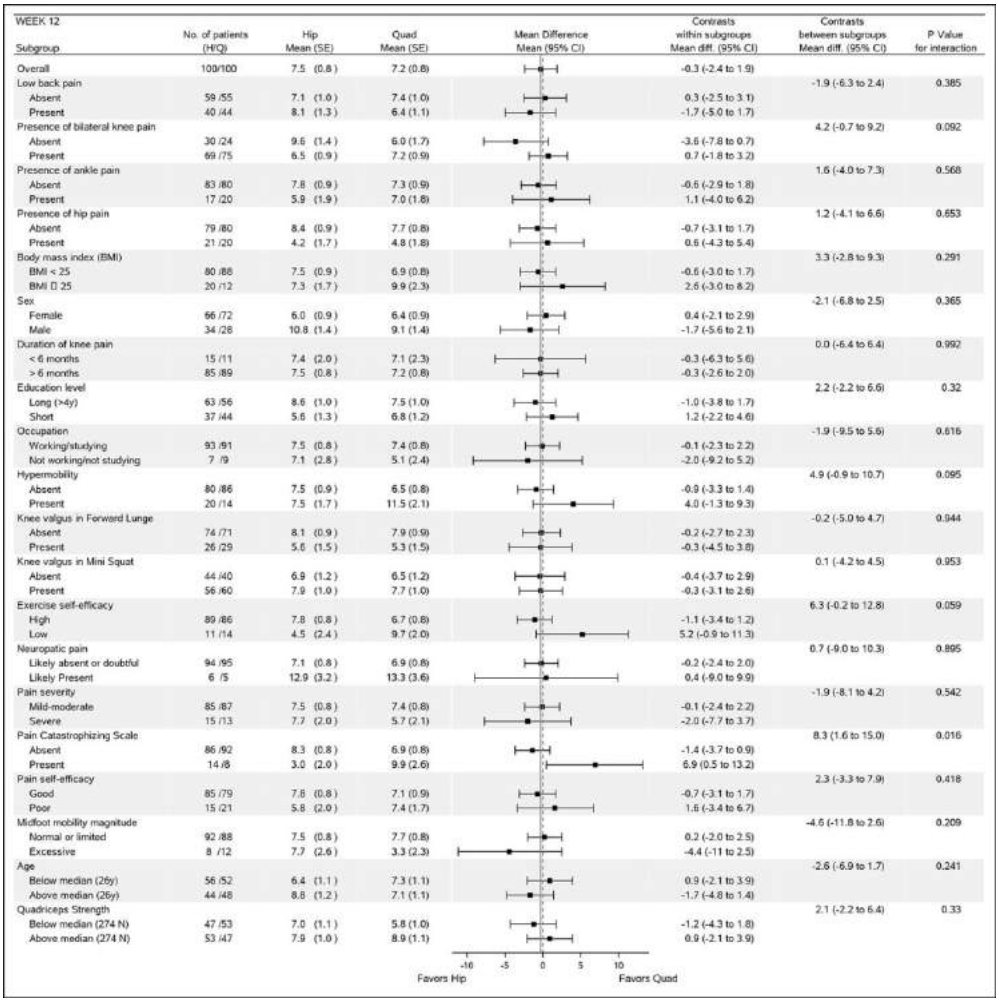


Figure 3. Forest plots of the Treatment effect (QE vs HE) across different subgroups at week 26 based on dichotomous baseline variables (effect modifiers) in the ITT population.

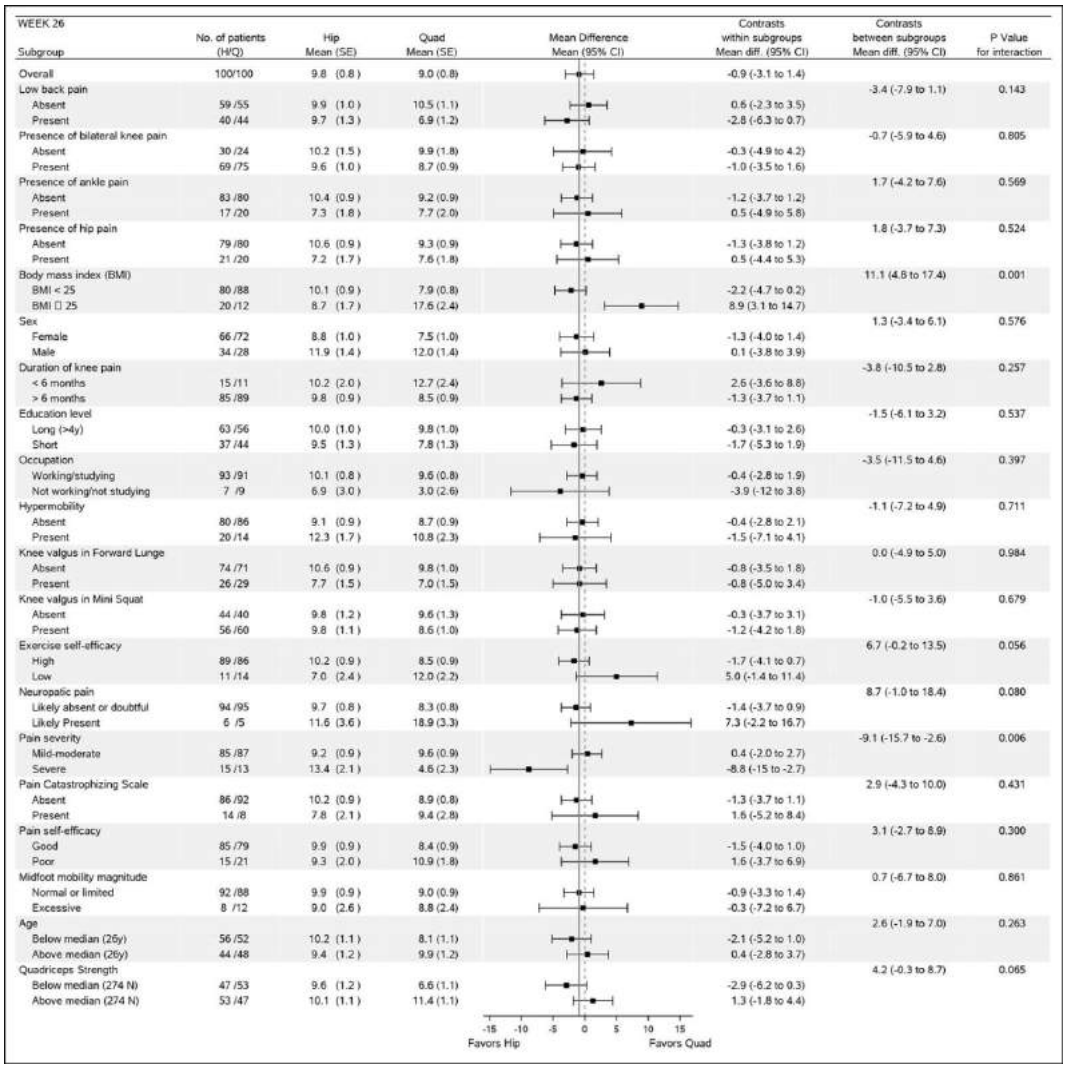


Table 1: *Dichotomous or dichotomized baseline characteristics used for subgrouping of participants.*

Variables	Description
Low back pain	Self-reported presence of low back pain during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", "Monthly" were defined as "Low back pain present" and the scores "No" and "Rarely" as "Low back pain not present".
Presence of bilateral knee pain	Self-reported presence of pain in the contralateral knee (not target knee) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", "Monthly" were defined as "Pain in the contralateral knee present" and the scores "No" and "Rarely" as "Pain in the contralateral knee not present".
Presence of ankle pain	Self-reported presence of ankle pain (one or both ankles) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", and "Monthly" were defined as "Ankle pain present" and the scores "No" and "Rarely" as "Ankle pain not present".
Presence of hip pain	Self-reported presence of hip pain (one or both hips) during the last 3 months. Scores were dichotomized: "Almost daily", "Several times during a week", "Weekly", and "Monthly" were defined as "Hip pain present" and the scores "No" and "Rarely" as "Hip pain not present".
Body mass index (BMI)	The participants' BMI is measured at baseline and dichotomized with a cut-off at 25 (BMI ≥ 25).
Sex	Female vs. male sex
Duration of knee pain	Self-reported duration of knee pain (present condition). We defined chronic pain as pain lasting for 6 months or longer.
Education level	Self-reported highest level of education: Scores were dichotomized: "Medium-term higher education (3-4 years)", and "Longer higher education (>4 years)" were defined as "Long education" and "Primary school", "Craftsman", "Highscool", and "Short higher education (<3 years)" as "Short education".
Occupation	Self-reported status of occupation and education. Scores were dichotomized: "Currently studying" and "Currently working" were defined as "Currently studying/working" and the score "Currently not working" as "Currently not working/not studying".
Hypermobility	Hypermobility assessed by the Beighton Score applying the revised criteria for the diagnosis of benign joint hypermobility syndrome. The Beighton score ranges from 0-9. Scores were dichotomized: scores 4-9 were defined as "Generalized joint hypermobility" and scores 0-3 as "Normal joint mobility". A score of 4 or more, is generally considered an indication of joint hypermobility (47, 48).
Quadriceps strength	Quadriceps strength is measured by handheld dynamometry. The muscle strength tests are conducted following validated testing protocols (49-51). As there is no established threshold available for sufficient or adequate muscle strength in PFP patients, measures were dichotomized according to the median strength (274 N) of all trial participants in the ITT population
Knee joint valgus malalignment during a forward lunge movement	Knee joint valgus malalignment assessed by clinical observation of the participant while he/she performs a forward lunge movement. The scores were dichotomized: The score "Definite valgus present" was defined as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".
Knee joint valgus malalignment during a single-leg squat movement	Knee joint valgus malalignment assessed by clinically observation of the participant while he/she performs a single-leg squat movement. The scores were dichotomized: The score "Definite valgus present" was defined as "Valgus malalignment" and the scores "No evidence of dynamic malalignment" and "definite varus present" as "No valgus malalignment".
Exercise self-efficacy	Self-reported exercise self-efficacy in relation to the two different exercise programs assessed by asking the participants to rate their confidence in performing the allocated exercise program on an 11-point (0-10) Likert scale with 0 representing "Not at all confident" and 10 representing "Completely confident". Scores were dichotomized: scores 6-10 were defined as "High self-efficacy" and the scores 0-5 as "Low self-efficacy".
Neuropatic pain	Presence of signs of neuropathic pain assessed by The painDETECT questionnaire (PDQ). A validated algorithm is used to calculate a total score ranging from -1 to 38. Scores were dichotomized: Scores ≥19 were defined as "Neuropathic pain component" and scores ≤18 as "No neuropathic pain component" as recommended by Freynhagen et al. (52).
Pain severity	Self-reported average pain during the past 4 weeks on a 0-10 Numeric Rating Scale. Scores were dichotomized: scores 0-6 were defined as "Mild or moderate pain" and the scores 7-10 as "Severe pain".
Pain Catastrophizing Scale	Presence of pain-related catastrophic thinking assessed by The Pain Catastrophizing Scale (PCS). Scores were dichotomized: scores > 30 were defined as "Pain catastrophizing" and scores 0 - 30 as "No pain catastrophizing". Previous studies have shown a cut-off of more than 30 points to be of clinical relevance (53, 54).

Variables	Description
Pain self-efficacy	Confidence in performing activities while in pain assessed by The pain self-efficacy questionnaire. Confidence in performing activities is rated on a 7-point (0-6) Likert scale with 0 representing not at all confident and 6 representing completely confident. A total score is calculated by summing the answers producing a score between 0 and 60. Scores were dichotomized: scores 0-39 was defined as "Poor pain self-efficacy" and the scores 40-60 as "Good self-efficacy". Scores around 40 (percentile = 50) are associated with return to work and maintenance of functional gains, whilst lower scores tend to predict less sustainable gains in injured workers (55).
Midfoot mobility magnitude	Midfoot mobility measured by the change in midfoot width from non-weight bearing to weight bearing. We defined a dichotomization of scores within the range 0-1.24 cm (0.92 cm +1*SD) as "Normal or limited midfoot mobility magnitude" and the scores above 1.25 cm as "Excessive midfoot mobility magnitude" for the females. For the males, we defined a dichotomization of scores within the range 0-1.36 cm (1.02 cm +1*SD) as "Normal or limited midfoot mobility magnitude and the scores above 1.37 cm as "Excessive midfoot mobility magnitude". Dichotomization is based on normative data for foot mobility (56).
Age	Age at inclusion in the trial. To split the trial population in two groups, we chose to categorize the participants based on the median age (26 years) of all trial participants in the ITT population.

Table 2. Baseline characteristics and summary of treatment effect modifiers

	Quadriceps exercise group (QE)	**Hip exercise group (HE)**
	n=100	**n=100**
Demographics		
Age, years	27.2 (6.7)	27.2 (6.3)
Duration of pain, months	52.8 (54.1)	47.3 (49.4)
Body mass, kg	67.6 (13.0)	68.2 (12.4)
Height, m	173.2 (10.7)	172.4 (8.4)
Body Mass Index, BMI (kg/m2)	22.4 (2.9)	22.8 (3.0)
Primary outcome		
AKPS questionnaire score (0-100)	73.3 (13.0)	74.2 (12.0)
Potential effect modifiers		
Female sex, n(%)	66 (66%)	72 (72%)
Above median age, n(%)	44 (44%)	48 (48%)
BMI≥ 25, n(%)	20 (20%)	12 (12%)
Chronic Knee pain (>6 months), n(%)	85 (85%)	89 (89%)
Presence of low back pain, n(%)	40 (40%)	44 (44%)
Presence of bilateral knee pain, n(%)	69 (69%)	75 (75%)
Presence of ankle pain , n(%)	17 (17%)	20 (20%)
Presence of hip pain, n(%)	21 (21%)	20 (20%)
Short education, n(%)	37 (37%)	44 (44%)
Currently not working/not studying, n(%)	7 (7%)	9 (9%)
Generalized joint hypermobility, n(%)	20 (20%)	14 (14%)
Knee joint valgus malalignment forward lunge, n(%)	26 (26%)	29 (29%)
Knee joint valgus malalignment single-leg squat, n(%)	56 (56%)	60 (60%)
Low self-efficacy, n(%)	11 (11%)	14 (14%)
Neuropathic pain component, n(%)	6 (6%)	5 (5%)
Severe pain, n(%)	15 (15%)	13 (13%)
Pain catastrophizing, n(%)	14 (14%)	8 (8%)
Poor pain self-efficacy, n(%)	15 (15%)	21 (21%)
Excessive mobility magnitude, n(%)	8 (8%)	12 (12%)
Low quadriceps strength, n(%)	47 (47%)	53 (53%)

REFERENCES

1. Rathleff MS, Rathleff CR, Olesen JL, Rasmussen S, Roos EM. Is Knee Pain During Adolescence a Self-limiting Condition? Prognosis of Patellofemoral Pain and Other Types of Knee Pain. Am J Sports Med. 2016;44(5):1165-71.
2. Rathleff CR, Baird WN, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Hip and knee strength is not affected in 12-16 year old adolescents with patellofemoral pain--a cross-sectional population-based study. PLoS One. 2013;8(11):e79153.
3. Roush JR, Curtis Bay R. Prevalence of anterior knee pain in 18-35 year-old females. Int J Sports Phys Ther. 2012;7(4):396-401.
4. Biber R, Gregory A. Overuse injuries in youth sports: is there such a thing as too much sports? Pediatr Ann. 2010;39(5):286-92.
5. Rathleff CR, Olesen JL, Roos EM, Rasmussen S, Rathleff MS. Half of 12-15-year-olds with knee pain still have pain after one year. Dan Med J. 2013;60(11):A4725.
6. Lankhorst NE, van Middelkoop M, Crossley KM, Bierma-Zeinstra SM, Oei EH, Vicenzino B, et al. Factors that predict a poor outcome 5-8 years after the diagnosis of patellofemoral pain: a multicentre observational analysis. Br J Sports Med. 2016;50(14):881-6.
7. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med. 2016.
8. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'Best Practice Guide to Conservative Management of Patellofemoral Pain': incorporating level 1 evidence with expert clinical reasoning. Br J Sports Med. 2015;49(14):923-34.
9. Collins NJ, Barton CJ, van Middelkoop M, Callaghan MJ, Rathleff MS, Vicenzino BT, et al. 2018 Consensus statement on exercise therapy and physical interventions (orthoses, taping and manual therapy) to treat patellofemoral pain: recommendations from the 5th International Patellofemoral Pain Research Retreat, Gold Coast, Australia, 2017. Br J Sports Med. 2018;52(18):1170-8.
10. van der Heijden RA, Lankhorst NE, van Linschoten R, Bierma-Zeinstra SM, van Middelkoop M. Exercise for treating patellofemoral pain syndrome. Cochrane Database Syst Rev. 2015;1:CD010387.
11. Winters M, Holden S, Lura CB, Welton NJ, Caldwell DM, Vicenzino BT, et al. Comparative effectiveness of treatments for patellofemoral pain: a living systematic review with network meta-analysis. Br J Sports Med. 2020.
12. Hott A, Brox JI, Pripp AH, Juel NG, Paulsen G, Liavaag S. Effectiveness of Isolated Hip Exercise, Knee Exercise, or Free Physical Activity for Patellofemoral Pain: A Randomized Controlled Trial. Am J Sports Med. 2019;47(6):1312-22.
13. Hansen RB, C.; Rathleff, MS.; Magnusson, SP.; Henriksen, M. Quadriceps or Hip Exercises for patellofemoral pain? A randomized controlled equivalence trial. Publication pending. 2022.
14. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. J Clin Epidemiol. 2021;134:174-7.
15. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther. 2009;89(7):698-704.
16. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley K, et al. Can we predict the outcome for people with patellofemoral pain? A systematic review on prognostic factors and treatment effect modifiers. Br J Sports Med. 2017;51(23):1650-60.

17. Lack S, Barton C, Vicenzino B, Morrissey D. Outcome predictors for conservative patellofemoral pain management: a systematic review and meta-analysis. Sports Med. 2014;44(12):1703-16.
18. Hott A, Brox JI, Pripp AH, Juel NG, Liavaag S. Predictors of Pain, Function, and Change in Patellofemoral Pain. Am J Sports Med. 2020;48(2):351-8.
19. Blond L, Hansen L. Patellofemoral pain syndrome in athletes: a 5.7-year retrospective follow-up study of 250 athletes. Acta Orthop Belg. 1998;64(4):393-400.
20. Collins NJ, Bierma-Zeinstra SM, Crossley KM, van Linschoten RL, Vicenzino B, van Middelkoop M. Prognostic factors for patellofemoral pain: a multicentre observational analysis. Br J Sports Med. 2013;47(4):227-33.
21. Collins NJ, Crossley KM, Darnell R, Vicenzino B. Predictors of short and long term outcome in patellofemoral pain syndrome: a prospective longitudinal study. BMC Musculoskelet Disord. 2010;11:11.
22. Natri A, Kannus P, Jarvinen M. Which factors predict the long-term outcome in chronic patellofemoral pain syndrome? A 7-yr prospective follow-up study. Med Sci Sports Exerc. 1998;30(11):1572-7.
23. Crossley KM, Stefanik JJ, Selfe J, Collins NJ, Davis IS, Powers CM, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br J Sports Med. 2016;50(14):839-43.
24. Lankhorst NE, van Middelkoop M, van Trier YD, van Linschoten R, Koes BW, Verhaar JA, et al. Can we predict which patients with patellofemoral pain are more likely to benefit from exercise therapy? A secondary exploratory analysis of a randomized controlled trial. J Orthop Sports Phys Ther. 2015;45(3):183-9.
25. Holden S, Rathleff MS, Jensen MB, Barton CJ. How can we implement exercise therapy for patellofemoral pain if we don't know what was prescribed? A systematic review. Br J Sports Med. 2018;52(6):385.
26. Toigo M, Boutellier U. New fundamental resistance exercise determinants of molecular and cellular muscle adaptations. Eur J Appl Physiol. 2006;97(6):643-63.
27. American College of Sports M. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. Med Sci Sports Exerc. 2009;41(3):687-708.
28. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy. 1993;9(2):159-63.
29. Crossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? Arch Phys Med Rehabil. 2004;85(5):815-22.
30. Sullivan MJL, Lynch ME, Clark AJ. Dimensions of catastrophic thinking associated with pain experience and disability in patients with neuropathic pain conditions. Pain. 2005;113(3):310-5.
31. Maclachlan LR, Collins NJ, Matthews MLG, Hodges PW, Vicenzino B. The psychological features of patellofemoral pain: a systematic review. Br J Sports Med. 2017;51(9):732-42.
32. Hott A, Pripp AH, Juel NG, Liavaag S, Brox JI. Self-efficacy and Emotional Distress in a Cohort With Patellofemoral Pain. Orthop J Sports Med. 2022;10(3):23259671221079672.
33. Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. Arthritis Rheum. 2000;43(5):995-1000.

34. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum. 2008;59(9):1207-13.
35. Hart HF, Barton CJ, Khan KM, Riel H, Crossley KM. Is body mass index associated with patellofemoral pain and patellofemoral osteoarthritis? A systematic review and meta-regression and analysis. Br J Sports Med. 2017;51(10):781-90.
36. Legha A, Burke DL, Foster NE, van der Windt DA, Quicke JG, Healey EL, et al. Do comorbidities predict pain and function in knee osteoarthritis following an exercise intervention, and do they moderate the effect of exercise? Analyses of data from three randomized controlled trials. Musculoskeletal Care. 2020;18(1):3-11.
37. Henriksen M, Nielsen SM, Christensen R, Kristensen LE, Bliddal H, Bartholdy C, et al. Who are likely to benefit from the Good Life with osteoArthritis in Denmark (GLAD) exercise and education program? An effect modifier analysis of a randomised controlled trial. Osteoarthritis Cartilage. 2022.
38. Staud R, Robinson ME, Price DD. Isometric exercise has opposite effects on central pain mechanisms in fibromyalgia patients compared to normal controls. Pain. 2005;118(1-2):176-84.
39. Lannersten L, Kosek E. Dysfunction of endogenous pain inhibition during exercise with painful muscles in patients with shoulder myalgia and fibromyalgia. Pain. 2010;151(1):77-86.
40. Burrows NJ, Booth J, Sturnieks DL, Barry BK. Acute resistance exercise and pressure pain sensitivity in knee osteoarthritis: a randomised crossover trial. Osteoarthritis Cartilage. 2014;22(3):407-14.
41. Vaegter HB. Exercising non-painful muscles can induce hypoalgesia in individuals with chronic pain. Scand J Pain. 2017;15:60-1.
42. Matthews M, Rathleff MS, Claus A, McPoil T, Nee R, Crossley KM, et al. Does foot mobility affect the outcome in the management of patellofemoral pain with foot orthoses versus hip exercises? A randomised clinical trial. Br J Sports Med. 2020;54(23):1416-22.
43. Lack S, Neal B, De Oliveira Silva D, Barton C. How to manage patellofemoral pain - Understanding the multifactorial nature and treatment options. Phys Ther Sport. 2018;32:155-66.
44. Barton CJ, Crossley KM. Sharing decision-making between patient and clinician: the next step in evidence-based practice for patellofemoral pain? Br J Sports Med. 2016;50(14):833-4.
45. Legare F, Stacey D, Turcotte S, Cossi MJ, Kryworuchko J, Graham ID, et al. Interventions for improving the adoption of shared decision making by healthcare professionals. Cochrane Database Syst Rev. 2014(9):CD006732.
46. Barton CJ, De Oliveira Silva D, Morton S, Collins NJ, Rathleff MS, Vicenzino B, et al. REPORT-PFP: a consensus from the International Patellofemoral Research Network to improve REPORTing of quantitative PatelloFemoral Pain studies. Br J Sports Med. 2021;55(20):1135-43.
47. Grahame R, Bird HA, Child A. The revised (Brighton 1998) criteria for the diagnosis of benign joint hypermobility syndrome (BJHS). J Rheumatol. 2000;27(7):1777-9.
48. Folci M, Capsoni F. Arthralgias, fatigue, paresthesias and visceral pain: can joint hypermobility solve the puzzle? A case report. BMC Musculoskelet Disord. 2016;17:58.
49. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. Scand J Med Sci Sports. 2010;20(3):493-501.
50. Romero-Franco N, Jimenez-Reyes P, Montano-Munuera JA. Validity and reliability of a low-cost digital dynamometer for measuring isometric strength of lower limb. J Sports Sci. 2017;35(22):2179-84.
51. Mentiplay BF, Perraton LG, Bower KJ, Adair B, Pua YH, Williams GP, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. PLoS One. 2015;10(10):e0140822.

52. Freynhagen R, Baron R, Gockel U, Tolle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006;22(10):1911-20.
53. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med. 2000;23(4):351-65.
54. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. J Behav Med. 1997;20(6):589-605.
55. Adams JH, Williams AC. What affects return to work for graduates of a pain management program with chronic upper limb pain? J Occup Rehabil. 2003;13(2):91-106.
56. McPoil TG, Vicenzino B, Cornwall MW, Collins N, Warren M. Reliability and normative values for the foot mobility magnitude: a composite measure of vertical and medial-lateral mobility of the midfoot. J Foot Ankle Res. 2009;2:6.

Study 3 – Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study

Note:

The supplementary file in the 'Results' section corresponds to the cross tabulated agreements found in Appendix 8

Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study

Rudi Hansen[1], Mathilde Lundgaard-Nielsen[1] and Marius Henriksen[2]

[1] Department of Physical and Occupational Therapy, Copenhagen University Hospital Bispebjerg and Frederiksberg, Copenhagen, Denmark
[2] The Parker Institute, Copenhagen University Hospital Bispebjerg and Frederiksberg, Copenhagen, Denmark

ABSTRACT

Background. Assessment of knee kinematics plays an important role in the clinical examination of patients with patellofemoral pain (PFP). There is evidence that visual assessments are reliable in healthy subjects, but there is a lack of evidence in injured populations. The purpose of this study was to examine the intra- and interrater agreement in the visual assessment of dynamic knee joint alignment in patients with PFP.

Methods. The study was a cross-sectional agreement study. Sixty participants (42 females) were included. We assessed the intra- and interrater agreement of two functional tests: The single leg squat (SLS) and the forward lunge (FL). One investigator scored the movement according to preset criteria while video recording the movement for retest. Moreover, the performance was scored by another investigator using the video recording. Agreement was assessed using weighted kappa statistics.

Results. The intrarater agreement ranged from moderate to good (Kappa 0.58 (FL) to 0.70 (SLS)) whereas the interrater agreement ranged from fair to moderate (Kappa 0.22 (SLS) to 0.50 (FL)).

Conclusion. The agreement within raters was better than between raters, which suggests that assessments should preferably be performed by the same tester in research and in a clinical setting, *e.g.*, to evaluate any treatment effect. We promote FL as a reliable clinical tool for evaluating dynamic knee alignment, since it shows equally good intra- and interrater agreement, and it is an inexpensive and easy method to use.

Submitted 17 June 2021
Accepted 2 September 2021
Published 19 October 2021

Corresponding author
Rudi Hansen, jth299@alumni.ku.dk

Academic editor
Guillermo Mendez-Rebolledo

Additional Information and Declarations can be found on page 10

DOI **10.7717/peerj.12203**

Subjects Anesthesiology and Pain Management, Orthopedics, Rheumatology, Biomechanics
Keywords Agreement, Knee Alignment, Patellofemoral Pain, PFP, Visual Assessment, Reliability, Knee Kinematics

BACKGROUND

Malalignment of the lower extremity have been linked to musculoskeletal problems, including patellofemoral pain (PFP) (*Myer et al., 2015*; *Willson, Binder-Macleod & Davis, 2008*; *Powers, 2003*; *Gwynne & Curran, 2018*). During movements such as squatting, individuals with PFP have demonstrated greater knee abduction excursion than controls (*Nakagawa et al., 2012*; *Willson & Davis, 2008*), and improvements in frontal and transverse plane pelvis and hip control have been linked to a reduction in pain (*Mascal, Landel & Powers, 2003*). Therefore, physiotherapists use visual assessment of dynamic

How to cite this article Hansen R, Lundgaard-Nielsen M, Henriksen M. 2021. Visual assessment of dynamic knee joint alignment in patients with patellofemoral pain: an agreement study. *PeerJ* **9**:e12203 http://doi.org/10.7717/peerj.12203

alignment in their clinical decision-making process when considering prescription of exercises (*Stephen et al., 2020*).

Dynamic alignment can be measured by clinical observation or biomechanical motion-analysis technology. Development of three-dimensional biomechanical analyses has made it possible to quantify and evaluate knee kinematics during functional tasks in a valid and reliable manner (*Mok, Petushek & Krosshaug, 2016*). However, this method is generally costly and too time consuming in the clinical setting. Two-dimensional measures, such as the frontal plane projection angle and visual assessments or ratings of lower extremity motion, have been suggested as more cost effective and acceptable alternatives to three-dimensional motion capture (*Willson & Davis, 2008*; *Harris-Hayes et al., 2014*).

In visual assessments of frontal plane knee motion during a single leg squat (SLS) and a forward lunge (FL), the reliability is reported to range from moderate to excellent in nonsymptomatic subjects (*Harris-Hayes et al., 2014*; *Weeks, Carty & Horan, 2012*; *Stensrud et al., 2011*; *Ageberg et al., 2010*; *Jones et al., 2014*; *Tate et al., 2015*). Measurement properties of a test instrument are likely to depend on the population studied, and there is a lack of evidence in injured populations (*Whatman, Toomey & Emery, 2019*). In PFP patients reliability measures have been evaluated using a 2-D video capture procedure that quantifies frontal plane knee alignment during single limb squats (*Gwynne & Curran, 2014*). The results of this study suggest that this method is reliable (ICC = 0.86) in PFP patients. Measures of visual assessment of lower extremity kinematics in PFP patients without the use of video analysis have been limited to a single study evaluating a lateral step-down task (*Piva et al., 2006*). This study included the evaluation of several aspects of movement quality (arm strategy, trunk movement, pelvis plane, knee position and balance) and scored each item according to a scale designed for the study to assess an overall movement quality. While this multimodal approach to evaluate movement quality may be more comprehensive, it is difficult to compare the results to other more commonly used measures of alignment and to extrapolate the findings to a clinical setting.

Because therapists use visual ratings to make clinical decisions, the reliability of these ratings needs to be considered. Our intention was to evaluate whether a simple visual rating of knee movement during two commonly used tests of dynamic alignment (the SLS and FL test) can be used reliably among PFP patients. The rating method used in this study resembles a clinical setting where the therapists do not have access to the equipment or time required for complex biomechanical analysis. The SLS and the FL were chosen because they are commonly used in clinical practice and have been reported in many previous studies investigating visual rating of lower extremity function (*Whatman, Hume & Hing, 2013*). The tests are less demanding than the commonly used jump tests; they also more closely resemble activities of daily life, such as stair ascending/descending, which may be more appropriate in a population of both sports-active and sedentary individuals.

The aims of this study were (a) to determine the intrarater agreement of a subjective visual assessment by an experienced sports physiotherapist in evaluating dynamic knee control during a SLS and a FL in a group of PFP patients, and (b) to determine the interrater agreement of the subjective assessment of dynamic knee control between two experienced sports physiotherapists.

MATERIALS & METHODS

Study design

The study was a cross-sectional agreement study. The reporting of the study follows the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (*Kottner et al., 2011*).

Participants

Participants were a subset of an RCT study aiming to compare the effectiveness of therapeutic hip and knee exercise for patients with PFP and to identify candidate patient characteristics that predict differential responses to the two exercise programs (clinicaltrials.gov identifier: NCT03069547). The study was pre-planned in the parent trial protocol in order to assess measurement properties of the dynamic knee joint alignment measures, that will be assessed as a potential patient characteristic associated with treatment response. The assessments were performed at baseline in the parent trial, *i.e.,* before randomization. As part of the baseline information gathered in the parent trial the participants answered the Anterior Knee Pain Score (a specific disability score for PFP patients ranging from 0 to 100 points, where higher scores indicate less disability (*Kujala et al., 1993*)), and self-reported pain during the past 4 weeks assessed on a 0 ('no pain') to 10 ('worst imaginable pain') Numeric Rating Scale. One participant failed to complete the Anterior Knee Pain Score and 3 participants failed to answer one of the 13 questions. The reported mean is calculated for 59 participants and missing data handled by imputing the mean of the participants for that particular item as recommended by *Hott et al. (2021)* and *Chavance (2004)*. Participants was recruited from the Institute of Sports Medicine Copenhagen (ISMC), Bispebjerg-Frederiksberg Hospital, Denmark. ISMC is a medical unit mainly for patients with injuries in the musculoskeletal system caused by participation in sports activities, and thus most participants are sports active. We aimed at including 60 participants in this sub-study, which gives 80% power to detect a kappa-coefficient of at least 0.5 that is statistically significantly different from 0 and corresponds to a moderate agreement.

Raters

Raters were RH (male) and MLN (female) who are both sports physiotherapists with 18 and 15 years of experience, respectively, in treating and assessing patients with musculoskeletal problems. Both raters use visual assessments as part of their daily clinical practice but have not used it for research purposes.

General procedures

Video was recorded using a tablet (Apple Ipad Air 2, frame rate: 30 frames per second) from an anterior view of participants performing the SLS and FL in the gym at the Department of Physical and Occupational Therapy at Bispebjerg and Frederiksberg Hospital. An investigator (RH) instructed the participants to perform the selected movement as described below. After the instruction had been understood, the participant performed the selected movement without rehearsal. If the participant lost his/her balance during the test, a new trial was performed. No discussion of the testing procedures or the classification criteria

occurred during the testing. The investigator filmed the participant and simultaneously scored the movement as observed clinically according to the criteria set below. The tablet was set up directly in front of participants, perpendicular to the frontal plane at a height of 100 cm and a distance of 3.5 m. The video captured the whole person. At least 1 week later, the investigator did another scoring based on the recorded video and another investigator (MLN) repeated the scoring independently. Three playbacks of the recorded video in real time were allowed for the intra- and interrater assessment.

Knee alignment during single leg squat

The SLS test has been described in several studies and the present procedure was a replication of comparable agreement studies (*Harris-Hayes et al., 2014*; *Ageberg et al., 2010*; *Nae et al., 2017*). From a position of single leg standing (painful knee), individuals performed a partial squat on one leg (hip and knee flexion) with the trunk maintained in an upright position, the contralateral hip in neutral and contralateral knee flexed. Individuals were instructed to perform the squat until they reached maximum ankle dorsiflexion without lifting their heels and then return to upright standing (Fig. 1A). The SLS was performed at participant-selected speed.

Knee alignment during forward lunge

The FL test was performed according to comparable agreement studies (*Nae et al., 2017*; *Alkjaer et al., 2002*). From a position of bilateral stance, individuals performed a forward step (painful knee), and continued the motion by flexing the front and back knee simultaneously (forward lunge).

Individuals were instructed to continue the lunge until reaching maximum dorsiflexion of the stance leg without lifting their heel and to push-off to upright position (Fig. 1B). The FL was performed at participant-selected speed. For both tests, dynamic valgus alignment was defined as an excessive medial movement of the knee as evidenced by an apparent increased frontal plane knee angle during the selected movement. Varus alignment was defined as an excessive lateral movement.

Scoring

The rater determined if a dynamic worsening of valgus or varus was present and scored the movement using the following categories:

−4 = severe valgus
−3 = moderate valgus
−2 = mild valgus
−1 = doubtful valgus
0 = no evidence of neither valgus nor varus
1 = doubtful varus
2 = mild varus
3 = moderate varus
4 = severe varus

We defined 'no evidence of neither valgus nor varus' as a neutral knee alignment, *i.e.*, knee flexion aligned with the 2nd toe. We considered 'doubtful' a just merely detectable

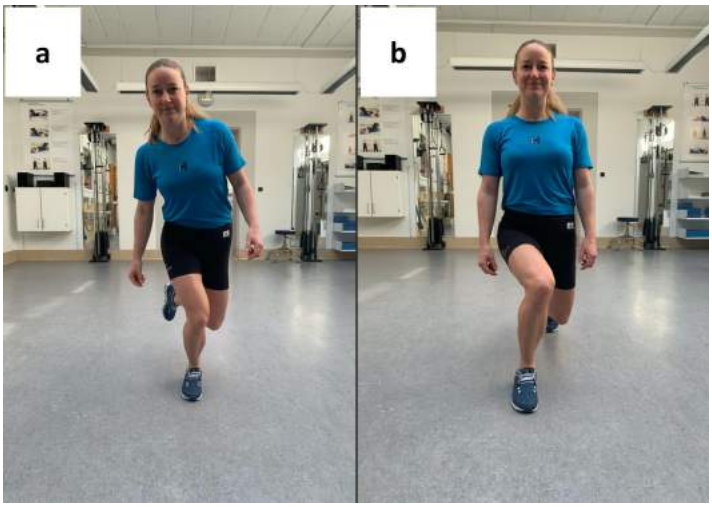


Figure 1 **Screenshots of video recordings in the assessment of a single leg squat (A) and a forward lunge (B).**

Full-size DOI: 10.7717/peerj.12203/fig-1

deviation from neutral alignment, while 'mild', 'moderate' and 'severe' was considered a definite deviation from neutral. The raters ranked the deviations 'mild', 'moderate' or 'severe' based on the experience of the raters in assessing PFP patients. We defined 'mild' as a slight deviation that might not be clinically relevant, 'moderate' as a modest and clinically relevant deviation from neutral, and 'severe' as a clinically relevant and severe collapse of the knee.

Similar ordinal and nominal scales have previously been used in the assessment of intra- and interrater agreement (*Weeks, Carty & Horan, 2012*; *Chmielewski et al., 2007*; *Junge et al., 2012*; *Trulsson, Garwicz & Ageberg, 2010*; *Whatman, Hing & Hume, 2012*). Visual assessments of dynamic knee joint alignment have been validated against a 'gold standard', *i.e.,* three-dimensional motion analysis, and the ability of visual assessments to identify 'true' malalignment is considered acceptable (*Ageberg et al., 2010*; *Whatman, Hume & Hing, 2013*).

Single leg squat and forward lunge classifications

An a priori categorical classification was made where the scores −4 to −2 were categorized as 'Definite valgus present', the scores −1, 0 and 1 were categorized as 'No definite evidence of dynamic malalignment' and the scores 2 to 4 were categorized as 'Definite varus present'. Conversion of scores into categorical variables is recommended in order to simplify the ratings (*Whatman, Hume & Hing, 2013*).

Statistical methods

Statistical analysis was completed using SAS (version 9.1 for Windows; SAS Institute Inc, Cary, NC). Descriptive statistics were calculated for demographics. Weighted kappa values with 95% CIs were used to examine the intratester and intertester reliability of the visual assessment. Cohens weighted kappa values with 95% CIs were used to examine the intratester and intertester reliability of the visual assessment. Cohens weighted kappa is broadly used and is a robust statistic useful for interrater and intrarater reliability testing (*McHugh, 2012*). Agreement was assessed using the classification for each movement test and for raw data (scores from −4 to 4). Interpretations of Kappa values were based on the guidelines adapted from *Landis & Koch (1977)*: <0.20: Poor agreement; 0.21–0.40: Fair agreement; 0.41–0.60: Moderate agreement; 0.61–0.80: Good agreement; 0.81–1.00: Very good agreement. We aimed at including 60 participants which gave 80% power to detect a kappa-coefficient of at least 0.5, which corresponds to moderate agreement.

Ethical considerations

Ethics approval was given by the Health Research Ethics Committee of the Capital Region of Denmark (De Videnskabsetiske Komitéer for Region Hovedstaden), protocol #H-16045755. Participants written informed consent were obtained prior to the start of the study.

RESULTS

The first sixty individuals with PFP who were included in the parent trial accepted the invitation to participate in this agreement study. Their characteristics are shown in Table 1. The dispersion of the data from the intra- and interrater assessments are presented in a heat map in Fig. 2 (SLS) and Fig. 3 (FL). The weighted kappa values for the classifications and the raw scores are shown in Tables 2 and 3. In summary, the intrarater agreement were statistically significantly different from 0 ($p < 0.0001$) and ranged from 0.58 to 0.70, *i.e.,* moderate to good agreement, whereas the interrater agreement ranged from 0.22 ($p = 0.08$) to 0.50 ($p < 0.0001$), *i.e.,* fair to moderate agreement. Interrater agreement was generally not as good as intrarater agreement (0.7 for SLS intrarater classification scores *vs.* 0.22 for interrater scores, and 0.58 for FL intrarater classification scores *vs.* 0.48 for interrater scores). The mean time from baseline to re-evaluation in the intrarater assessment was 29.1 days (SD 14.8). The cross tabulated agreements in the classifications and raw scores are provided in the Supplementary File.

DISCUSSION

Principal findings

The aim of this study was to determine the intra- and interrater agreement between two experienced physiotherapists when visually assessing the dynamic knee alignment during an SLS and an FL in a population of PFP patients. The most important finding was that the visual assessments of dynamic alignment during SLS and FL can be done reliably when the assessment is repeated by the same rater. Moreover, 'moderate' levels of agreements

Table 1 **Descriptive characteristics of the participants ($n = 60$).**

Characteristics	Mean (SD)
Age (yrs)	27.2 (6.2)
Females (n (%))	42 (70%)
Height (cm)	172.1 (8.6)
Weight (kg)	66.6 (10.6)
Body mass index	22.4 (2.8)
Duration of symptoms (months)	34 (34)
Anterior Knee Pain Score (0–100 points)	30.20 (5.15)
Average pain during previous 4 weeks (NRS* 0–10)	3.73 (2.17)

Notes.
*Numeric rating scale.

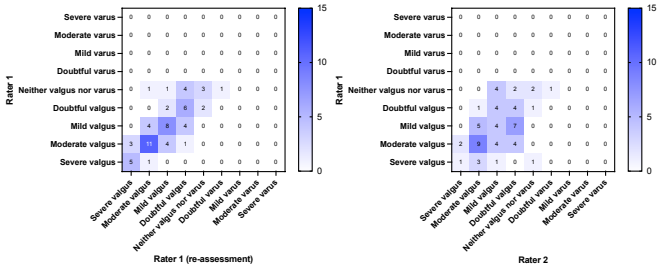


Figure 2 **Heatmap of agreement matrix showing the dispersion of the intrarater (left) and interrater agreement (right) for the single leg squat.** The brightness of the blue color indicates the number of rating combinations with darker colors representing higher numbers as shown in the individual squares and in the key bar.

Full-size DOI: 10.7717/peerj.12203/fig-2

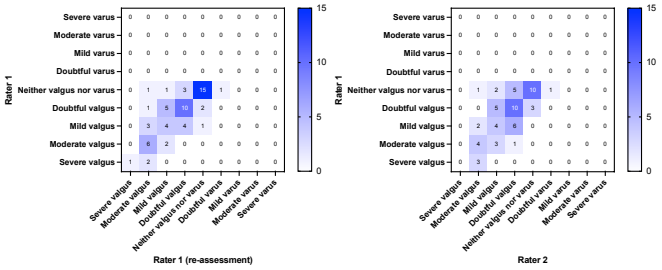


Figure 3 **Heatmap of agreement matrix showing the dispersion of the intrarater (left) and interrater agreement (right) for the forward lunge.** The brightness of the blue color indicates the number of rating combinations with darker colors representing higher numbers as shown in the individual squares and in the key bar.

Full-size DOI: 10.7717/peerj.12203/fig-3

Table 2 Intra- and interrater agreement of single leg squat.

	Weighted kappa value	95% CI	*p*-value
Intrarater classification	0.70	0.51–0.89	<.0001
Intrarater raw data	0.65	0.54–0.76	<.0001
Interrater classification	0.22	−0.03–0.48	0.08
Interrater raw data	0.32	0.16–0.48	0.05

Table 3 Intra- and interrater agreement of forward lunge.

	Weighted kappa value	95% CI	*p*-value
Intrarater classification	0.58	0.37–0.79	<.0001
Intrarater raw data	0.65	0.53–0.78	<.0001
Interrater classification	0.48	0.25–0.7	0.0002
Interrater raw data	0.50	0.36–0.64	<.0001

are seen when two experienced raters assess the FL, while the interrater agreement is only 'fair' when assessing the SLS.

Comparison with previous studies

The intrarater agreements for the SLS in the current study compare to the results of a review on the interrater and intrarater agreement of observation-based assessment of the SLS including studies on both healthy subjects and subjects with lower extremity disorders (*Ressman, Grooten & Barr, 2019*). In that study, the pooled results of Kappa showed a 'substantial' agreement for intrarater agreement (Kappa value 0.68 (95% CI [0.60–0.74]). Moreover, the review found a 'moderate' agreement for interrater reliability of the SLS (Kappa value 0.58 (95% CI [0.50–0.65]), which is somewhat higher than in present study. In the present study no efforts were made to synchronize assessors by mutual training sessions or by operationalizing the measurements. Visual assessments were therefore entirely based on the experience of the assessors. This might explain the discrepancy with the systematic review and the relatively low interrater agreements seen.

Clinical implications

The clinical implications of the results of our study are, that visual assessment of frontal plane knee kinematics during a FL can be done reliably by experienced testers, whereas SLS should preferably be re-evaluated by the same tester. Forward lunge is therefore a reliable clinical tool for evaluating knee alignment. Furthermore, it is an inexpensive and easy method to use, making it ideal for the clinical setting. However, the clinical validity, relevance, and prognostic value still need to be established.

Strengths and limitations

This study has some strengths and limitations. Firstly, by using an objective measurement tool (the tablet) we made sure that assessments were based on the same movement, and without verbal or non-verbal interaction between raters. Using the tablet, on the other hand, implies an inherent limitation by not taking the variability of the patients' performances

into account. It was out of the scope of this study to include a clinical re-evaluation of the participants and we thereby accept the exclusion of data on within subject variability. On the other hand, this enables a focused analysis of the agreement within and between raters.

Another strength is, that we have included a relatively large sample of PFP patients, which makes the results transferable to other clinical and experimental settings. It should be noted, though, that assessors were experienced, and our results may only be transferable when assessments are made by equally experienced assessors.

In our assessment of movement quality, we only included the knee excursion during the movement (dynamic knee valgus or varus). Rating dynamic knee alignment per se is not an exhaustive evaluation of movement quality and is merely one aspect of the full range clinical examinations. It is, however, considered a good indicator of movement quality (*Sahrmann, Azevedo & Dillen, 2017*) with a knee-medial-to-foot position often considered less optimal, indicating poor postural orientation (*Ageberg et al., 2010*; *Ortqvist et al., 2011*). A limitation of assessing movement quality without three-dimensional analyses is, that we lack information on movement components like transversal and sagittal plane control of body segments. However, we aimed solely at assessing the reliability of the test assessments and not on the validity of the actual assessments. Furthermore, we did not control for the speed or acceleration of the movement. We intentionally omitted the control in order to comply with our intents of resembling a clinical setup. We acknowledge the limitation of not including the control of speed and acceleration of the execution of the exercises and that speed and acceleration may have influenced the knee excursion.

Various scoring systems have been used to assess dynamic knee joint alignment in the literature (*Ressman, Grooten & Barr, 2019*). The scoring system used in this study resembles previous used systems (*Chmielewski et al., 2007*; *Whatman, Hing & Hume, 2012*). *Chmielewski et al., (2007)* used the terms "no deviation from neutral alignment", "small-magnitude or barely observable movement out of a neutral position", "moderate-magnitude or marked movement out of a neutral position, and "excessive or severe magnitude of movement out of a neutral position", to assess the degree of knee excursion during a unilateral squat and lateral step-down task. We made the scoring two-tailed (varus and valgus) to be more specific in the direction of the knee in the movement. The use of different scoring systems in the literature, makes it difficult to compare and pool the results of agreement studies. Future studies should consider standardizing the scoring for the benefit of research in reliability and agreement of knee kinematics.

No varus knees (>1 on the scoring scale) were found in the population. This means that when scores were converted into classifications there were only two viable options so it is difficult to tell if the results would be similar if individuals with varus were included in the analysis. The narrow range of scores is probably linked to the population, indicating that PFP patients are more prone to a valgus knee alignment.

Comparing real time scoring to retrospective scoring may have impacted our results. We chose the real time visual assessment in order to resemble a clinical setup and accepted the potential bias of re-evaluating the movement on a screen. We argue, however, that since the assessment was only in the frontal plane, the risk of bias in the subsequent re-rating on

a screen was small. Further, the re-assessment of the video recordings did not include slow motion, which makes the validity of the results representative of clinical practice.

CONCLUSIONS

Visual assessments of dynamic knee joint alignment during a FL and a SLS performed by patients with PFP can be done reliably when an experienced rater repeats the assessment. Two experienced physiotherapists agree moderately when assessing dynamic alignment during FL, but only 'fair' when assessing dynamic alignment during SLS. The agreement within raters is better than between raters, which suggests that assessments should preferably be performed by the same tester in research and in a clinical setting, *e.g.*, to evaluate any treatment effect. We suggest the FL as a reliable clinical tool for evaluating knee alignment in the clinical setting, since it shows acceptable intra- and interrater agreement, and it is an inexpensive and easy test to use.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Rudi Hansen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mathilde Lundgaard-Nielsen performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Marius Henriksen conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Ethics approval was given by the Health Research Ethics Committee of the Capital Region of Denmark, protocol #H-16045755.

Data Availability

The following information was supplied regarding data availability:

The cross tabulated agreements in the classifications and raw scores are available in the Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12203#supplemental-information.

REFERENCES

Ageberg E, Bennell KL, Hunt MA, Simic M, Roos EM, Creaby MW. 2010. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. *BMC Musculoskeletal Disorders* **11**:265 DOI 10.1186/1471-2474-11-265.

Alkjaer T, Simonsen EB, Peter Magnusson SP, Aagaard H, Dyhre-Poulsen P. 2002. Differences in the movement pattern of a forward lunge in two types of anterior cruciate ligament deficient patients: copers and non-copers. *Clinical Biomechanics* **17(8)**:586–593 DOI 10.1016/S0268-0033(02)00098-0.

Chavance M. 2004. Handling missing items in quality of life studies. *Communications in Statistics - Theory and Methods* **33(6)**:1371–1383 DOI 10.1081/STA-120030155.

Chmielewski TL, Hodges MJ, Horodyski M, Bishop MD, Conrad BP, Tillman SM. 2007. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. *Journal of Orthopaedic and Sports Physical Therapy* **37(3)**:122–129 DOI 10.2519/jospt.2007.2457.

Gwynne CR, Curran SA. 2014. Quantifying frontal plane knee motion during single limb squats: reliability and validity of 2-dimensional measures. *International Journal of Sports Physical Therapy* **9(7)**:898–906.

Gwynne CR, Curran SA. 2018. Two-dimensional frontal plane projection angle can identify subgroups of patellofemoral pain patients who demonstrate dynamic knee valgus. *Clinical Biomechanics* **58**:44–48 DOI 10.1016/j.clinbiomech.2018.06.021.

Harris-Hayes M, Steger-May K, Koh C, Royer NK, Graci V, Salsich GB. 2014. Classification of lower extremity movement patterns based on visual assessment: reliability and correlation with 2-dimensional video analysis. *Journal of Athletic Training* **49(3)**:304–310 DOI 10.4085/1062-6050-49.2.21.

Hott A, Liavaag S, Juel NG, Brox JI, Ekeberg OM. 2021. The reliability validity interpretability and responsiveness of the Norwegian version of the Anterior Knee Pain Scale in patellofemoral pain. *Disability and Rehabilitation* **43(11)**:1605–1614 DOI 10.1080/09638288.2019.1671499.

Jones D, Tillman SM, Tofte K, Mizner RL, Greenberg S, Moser MW, Chmielewski TL. 2014. Observational ratings of frontal plane knee position are related to the frontal plane projection angle but not the knee abduction angle during a step-down task. *Journal of Orthopaedic and Sports Physical Therapy* **44(12)**:973–978 DOI 10.2519/jospt.2014.4960.

Junge T, Balsnes S, Runge L, Juul-Kristensen B, Wedderkopp N. 2012. Single leg mini squat: an inter-tester reproducibility study of children in the age of 9-10 and 12-14 years presented by various methods of kappa calculation. *BMC Musculoskeletal Disorders* **13**:203 DOI 10.1186/1471-2474-13-203.

Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International Journal of Nursing Studies* **48(6)**:661–671 DOI 10.1016/j.ijnurstu.2011.01.016.

Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. 1993. Scoring of patellofemoral disorders. *Arthroscopy* **9(2)**:159–163 DOI 10.1016/S0749-8063(05)80366-4.

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics* **33(1)**:159–174 DOI 10.2307/2529310.

Mascal CL, Landel R, Powers C. 2003. Management of patellofemoral pain targeting hip, pelvis, and trunk muscle function: 2 case reports. *Journal of Orthopaedic and Sports Physical Therapy* **33(11)**:647–660 DOI 10.2519/jospt.2003.33.11.647.

McHugh ML. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* **22(3)**:276–282.

Mok KM, Petushek E, Krosshaug T. 2016. Reliability of knee biomechanics during a vertical drop jump in elite female athletes. *Gait & Posture* **46**:173–178 DOI 10.1016/j.gaitpost.2016.03.003.

Myer GD, Ford KR, Di Stasi SL, Foss KD, Micheli LJ, Hewett TE. 2015. High knee abduction moments are common risk factors for patellofemoral pain (PFP) and anterior cruciate ligament (ACL) injury in girls: is PFP itself a predictor for subsequent ACL injury? *British Journal of Sports Medicine* **49(2)**:118–122 DOI 10.1136/bjsports-2013-092536.

Nae J, Creaby MW, Nilsson G, Crossley KM, Ageberg E. 2017. Measurement properties of a test battery to assess postural orientation during functional tasks in patients undergoing anterior cruciate ligament injury rehabilitation. *Journal of Orthopaedic and Sports Physical Therapy* **47(11)**:863–873.

Nakagawa TH, Moriya ET, Maciel CD, Serrao FV. 2012. Trunk, pelvis, hip, and knee kinematics, hip strength, and gluteal muscle activation during a single-leg squat in males and females with and without patellofemoral pain syndrome. *Journal of Orthopaedic and Sports Physical Therapy* **42(6)**:491–501 DOI 10.2519/jospt.2012.3987.

Ortqvist M, Mostrom EB, Roos EM, Lundell P, Janarv PM, Werner S, Broström EW. 2011. Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children. *Knee Surgery, Sports Traumatology, Arthroscopy* **19(12)**:2060–2066 DOI 10.1007/s00167-011-1542-9.

Piva SR, Fitzgerald K, Irrgang JJ, Jones S, Hando BR, Browder DA, Childs JD. 2006. Reliability of measures of impairments associated with patellofemoral pain syndrome. *BMC Musculoskeletal Disorders* **7**:33 DOI 10.1186/1471-2474-7-33.

Powers CM. 2003. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. *Journal of Orthopaedic and Sports Physical Therapy* **33(11)**:639–646 DOI 10.2519/jospt.2003.33.11.639.

Ressman J, Grooten WJA, Rasmussen Barr E. 2019. Visual assessment of movement quality in the single leg squat test: a review and meta-analysis of inter-rater

and intrarater reliability. *BMJ Open Sport & Exercise Medicine* **5(1)**:e000541 DOI 10.1136/bmjsem-2019-000541.

Sahrmann S, Azevedo DC, Dillen LV. 2017. Diagnosis and treatment of movement system impairment syndromes. *Brazilian Journal of Physical Therapy* **21(6)**:391–399 DOI 10.1016/j.bjpt.2017.08.001.

Stensrud S, Myklebust G, Kristianslund E, Bahr R, Krosshaug T. 2011. Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players. *British Journal of Sports Medicine* **45(7)**:589–595 DOI 10.1136/bjsm.2010.078287.

Stephen J, Ephgrave C, Ball S, Church S. 2020. Current concepts in the management of patellofemoral pain—the role of alignment. *Knee* **27(2)**:280–286 DOI 10.1016/j.knee.2019.12.006.

Tate J, True H, Dale B, Baker C. 2015. Expert versus novice interrater and intrarater reliability of the frontal plane projection angle during a single-leg squat. *International Journal of Athletic Therapy and Training* **20(4)**:23–27 DOI 10.1123/ijatt.2014-0116.

Trulsson A, Garwicz M, Ageberg E. 2010. Postural orientation in subjects with anterior cruciate ligament injury: development and first evaluation of a new observational test battery. *Knee Surg Sports Traumatol Arthrosc* **18(6)**:814–823 DOI 10.1007/s00167-009-0959-x.

Weeks BK, Carty CP, Horan SA. 2012. Kinematic predictors of single-leg squat performance: a comparison of experienced physiotherapists and student physiotherapists. *BMC Musculoskeletal Disorders* **13**:207 DOI 10.1186/1471-2474-13-207.

Whatman C, Hing W, Hume P. 2012. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Physical Therapy in Sport* **13(2)**:87–96 DOI 10.1016/j.ptsp.2011.07.001.

Whatman C, Hume P, Hing W. 2013. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. *Physical Therapy in Sport* **14(3)**:168–174 DOI 10.1016/j.ptsp.2012.07.001.

Whatman C, Toomey C, Emery C. 2019. Visual rating of movement quality in individuals with and without a history of intra-articular knee injury. *Physiotherapy Theory and Practice* 1–7 DOI 10.1080/09593985.2019.1703229.

Willson JD, Binder-Macleod S, Davis IS. 2008. Lower extremity jumping mechanics of female athletes with and without patellofemoral pain before and after exertion. *American Journal of Sports Medicine* **36(8)**:1587–1596 DOI 10.1177/0363546508315592.

Willson JD, Davis IS. 2008. Utility of the frontal plane projection angle in females with patellofemoral pain. *Journal of Orthopaedic and Sports Physical Therapy* **38(10)**:606–615 DOI 10.2519/jospt.2008.2706.