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

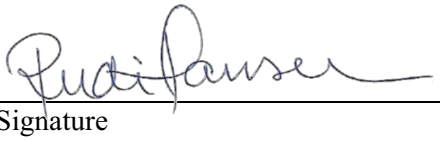
Protocol version	Updated SAP version	Section Number Changed	Description of and reason for change	Date changed

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

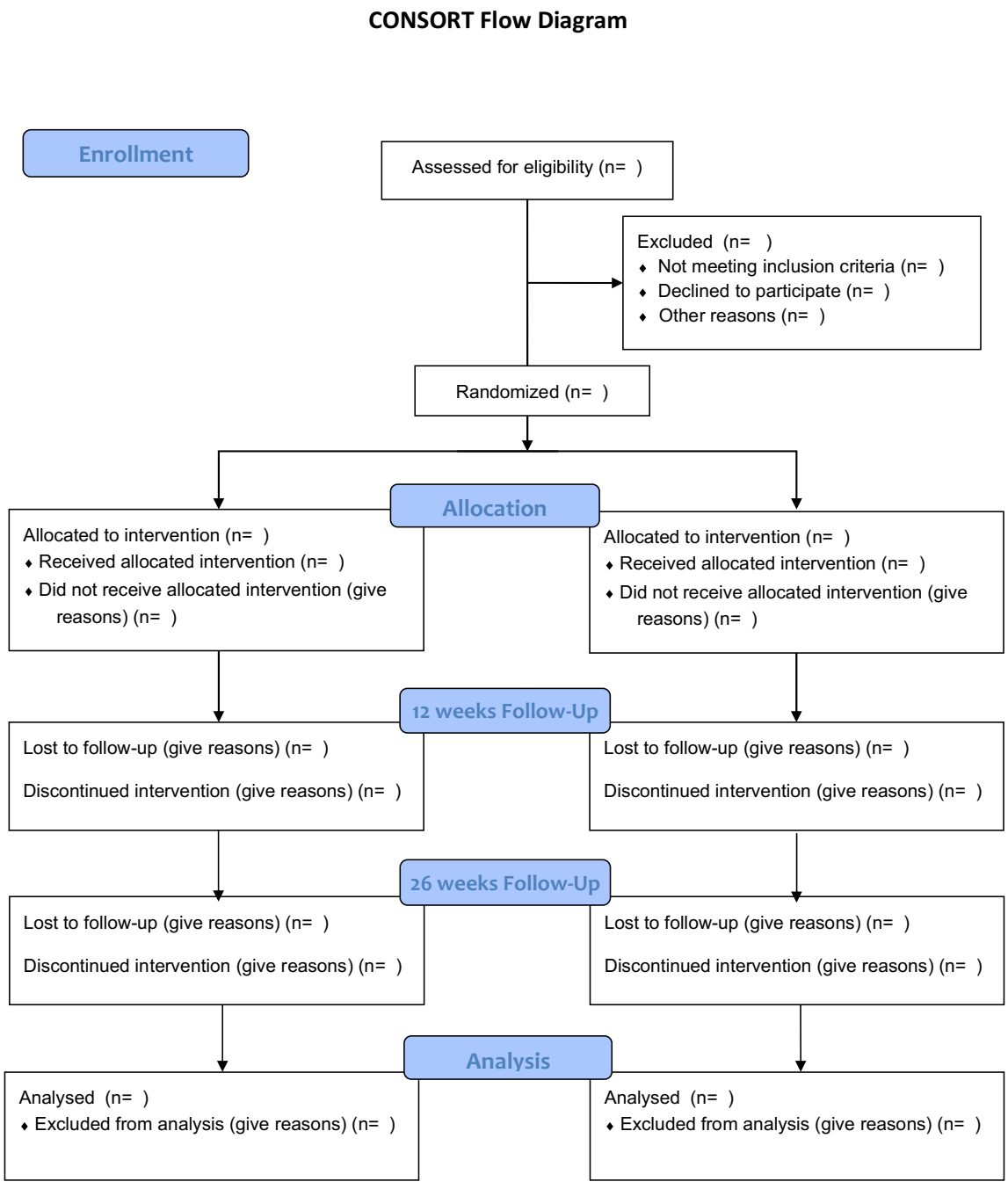
The following details in this SAP represents deviations from trial protocol version 1.5

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

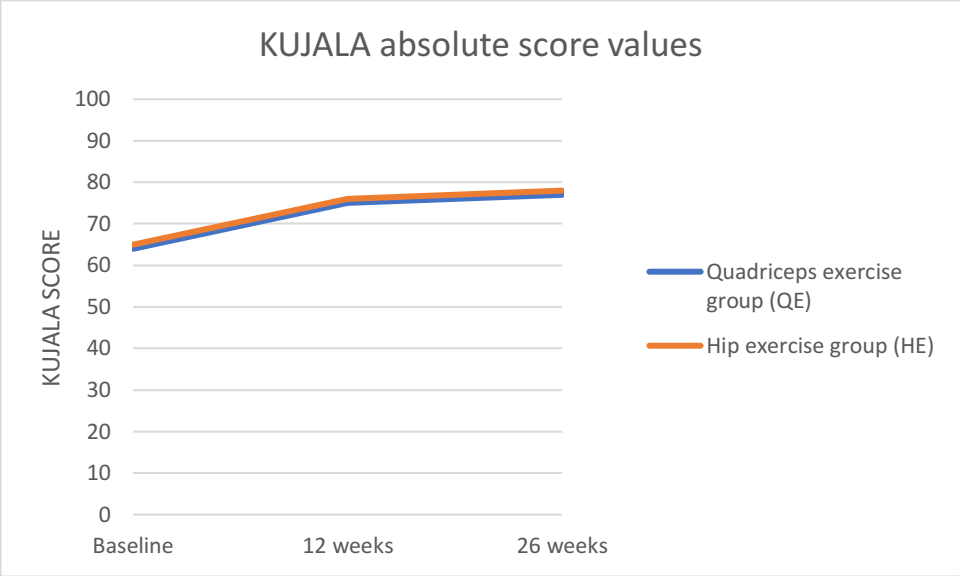
Table X: Primary and Secondary Outcomes at week 12 in the ITT population. CI denotes 95% confidence interval. Based on repeated measures mixed linear models, where missing data is modelled implicitly.

	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.