STATISTICAL ANALYSIS PLAN (SAP)

Title

First-line treatment for femoroacetabular impingement syndrome: study protocol for a multicenter randomized controlled trial comparing a 6-month supervised strength exercise intervention to usual care on hip-related quality of life. ("The Better Hip Trial").

ClinicalTrials.gov identifier: NCT05927935

Authors: Frederik Nicolai Foldager[1,2], Signe Kierkegaard-Brøchner[2,3], Joanne L Kemp[4], Maurits van Tulder[5], Ulrik Dalgas[6], Inger Mechlenburg[1,3,6] and Bo Martin Bibby[7]

Affiliations:

[1]Department of Orthopaedic Surgery, Aarhus University Hospital, Denmark

[2]Department of Physio and Occupational Therapy and Orthopaedic Surgery, Horsens Regional Hospital, Denmark

[3]Department of Clinical Medicine, Aarhus University, Denmark.

[4]La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Australia

[5]Faculty of Behavioural and Movement Sciences, Vrije University, Amsterdam

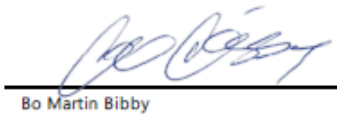
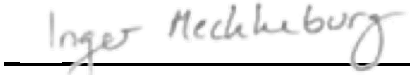
[6]Exercise Biology, Department of Public Health, Aarhus University, Denmark

[7]Department of Biostatistics, Institute of Public Health, Aarhus University, Aarhus, Denmark

Statistical Analysis Plan, Version: Original, 28-05-2025

Corresponding author: Frederik Nicolai Foldager: frederikfoldager@clin.au.dk

SIGNATURE PAGE

Role	Name	Date
Person writing the Statistical Analysis Plan (SAP)	Frederik Nicolai Foldager	28-05-2025
Signature:  Frederik Nicolai Foldager MSc, PT		
Senior statistician responsible	Bo Martin Bibby	28-05-2025
Signature:  Bo Martin Bibby Associate Professor		
Chief investigator / Clinical lead	Inger Mechlenburg	28-05-2025
Signature:  Inger Mechlenburg Professor		
Health economic evaluation lead	Maurits Van Tulder	28-05-2025
Signature:  Maurits Van Tulder Professor		

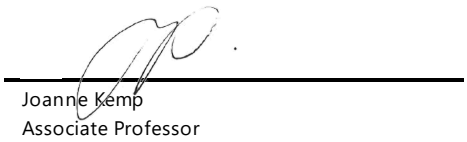
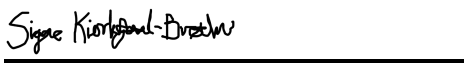
Chief Investigator / Clinical Lead in Australia	Joanne Kemp	28-05-2025
Signature:  Joanne Kemp Associate Professor		
Investigator / Clinical Lead in Horsens	Signe Kierkegaard-Brøchner	28-05-2025
Signature:  Signe Kierkegaard-Brøchner PostDoc		
Secondary analysis: Exercise adherence and volume lead	Ulrik Dalgas	28-05-2025
Signature:  Ulrik Dalgas Professor		

Table of content

INTRODUCTION:

Aim and hypothesis

We will conduct a randomized controlled trial (RCT) with economic evaluation investigating the clinical and cost-effectiveness of a physiotherapist-led strength exercise intervention compared to usual care as first-line treatment in patients with femoroacetabular impingement syndrome (FAIS). Furthermore, we will explore how exercise adherence and volume of a physiotherapist-led strength exercise intervention are associated with outcomes. We hypothesized that (i) 6-months of physiotherapist-led strength exercise intervention is superior (i.e. ≥ 6 points difference on the International Hip and Outcome Tool 33; iHOT-33), to usual care in improving hip related quality of life (QoL) in patients with FAIS after a 6 month intervention, (ii) 6-months of physiotherapist-led strength exercise intervention is cost-effective compared to usual first-line care at 12-months follow-up in patients with FAIS, and (iii) high exercise session adherence and high volume of exercise will be superior to low exercise adherence and low volume in moderating objectively- and patient-reported outcome measures in patients with FAIS.

STUDY METHODS

Trial design

The Better Hip trial is a multicenter, stratified (by hospital site), randomized (allocation 1:1), controlled, parallel-group, assessor-blinded, superiority trial conducted in Denmark (Aarhus, Horsens, Hvidovre, Aalborg, and Odense) and Australia (Melbourne).

Sample size calculation

The sample size calculation is based on a clinical superiority calculation using data from a pilot RCT by Kemp et al.[1] Kemp et al. found a mean change of 27 (SD 26) points on the iHOT-33 in their strength exercise group (e.g., intervention group) and a mean change of 11 (SD 8) points in their stretching group (e.g., control group) after 12 weeks of intervention (between-group difference of 16 [95% CI -9; 38] points). The superiority margin is a 6 points MCID in the between-group change on the iHOT-33.[2] The estimated mean difference of 16 points, a superiority margin of 6 points[2], a SD of 19.2[1], a power of 80% and an alpha level of 5% results in a total sample size of n=94 (n=47 in each group) After allowing a dropout-rate of up to 28% our required sample size will be n=120 (n=60 in each group).

Randomization

After baseline assessment, patients will be randomly assigned to either a supervised strength exercise intervention or usual care with a 1:1 allocation as per a computer-generated randomization schedule, stratified by hospital site (Aarhus, Odense, Aalborg, Horsens, Hvidovre, or Melbourne).

Statistical interim analyses and stopping guidance

No interim analysis or stopping guidance is planned as the interventions, supervised strength exercise, and usual care (e.g., advice to remain physically active following World Health Organization guidelines) are known to have minimal risks.

Timing of outcome assessments

Outcome assessments are planned at baseline and 6 months follow-up for performance-based physical functioning tests (e.g., maximal muscle strength in the one-repetition maximum unilateral leg press, single leg jump for distance, and one leg rise test). Outcome assessments are planned at baseline, 3, 6, and 12-month follow-ups for patient-reported outcome measures and cost. See Table 1 in the study protocol.[3]

STATISTICAL PRINCIPLES

Confidence intervals and p-values

Results will be expressed as between-group differences in means with 95% confidence intervals.

TRIAL POPULATION

All potential participants at the hospital sites will be screened for eligibility following the inclusion and exclusion criteria.

Inclusion criteria have previously been used for diagnostic inclusion in the PhysioFIRST trial[4]: (1) Activity- or position-related pain lasting ≥ 3 months; (2) Positive Flexion-Adduction-Internal Rotation (FADIR) test; (3) Cam-type FAIS; x-ray alpha angle > 60 degrees or 4) Pincer-type FAIS; lateral center edge angle (LCEA) > 39 degrees or cross-over sign or 5) Mixed-type FAIS; a combination of cam- and pincer-type impingement defined in inclusion criteria 3 and 4; (6) When asked, the participant states that they are

motivated to exercise twice a week for six months; (7) 18-50 years old; and (7) Body Mass Index (BMI) score < 35 (kg/m^2). Exclusion criteria were: (1) Involvement in supervised strengthening exercises targeting the hip muscles in the last 3 months before inclusion consisting of ≥ 6 supervised physiotherapy sessions; (2) Previous hip surgery or other major hip injuries in the index hip; (3) Evidence of pre-existing osteoarthritis, defined as Tönnis grade ≥ 2 or Kellgreen-Lawrence ≥ 2; (4) Evidence of pre-existing osteoarthritis, defined as lateral joint space width < 3 mm; (5) Hip dysplasia, defined as a CE-angle < 25° and an acetabular index (AI) angle > 10°; (6) Comorbidities or other problems considered to prevent participation in exercise and (7) Unable to communicate in the respective languages (Danish or English) of the participating countries. Eligibility will be limited to the hip with the highest level of symptoms at baseline.

Results will be presented in a CONSORT flowchart (See Figure 1) displaying the total number of participants who were: (1) screened, (2) excluded (with reasons), (3) randomized, (4) received allocated treatment, (5) discontinued intervention (with reasons), (6) lost to follow-up (with time and reason), (7) included in ITT analysis and (8) included in the per-protocol analysis.

Withdrawal/follow-up

Participants deciding to withdraw from the intervention will be asked to complete outcome assessments even though they stop attending exercise sessions or receive hip surgery.

Baseline participant characteristics

The following demographic and descriptive data will be obtained at baseline and presented by randomization group: biological sex (female/male/other), gender women/girl/men/boy/trans/non-binary/other), age (years), height (cm), weight (kg), body mass index (kg/m^2), index hip (left/right), hip symptom duration (months), previous hip surgery (right/left/none), previous treatment due to hip symptoms (no/yes; type of treatment), bilateral symptoms (no/yes), educational level, employment status, cohabitation status, smoking status, medicine consumption (no/yes; type of medicine; frequency) and comorbidities (no/yes; diagnosis/diagnoses). Baseline participant characteristics will be presented in Table 1.

ANALYSIS

Outcomes:

Primary outcome

International Hip and Outcome Tool 33 (iHOT-33).

The primary outcome is the change in the hip-related quality of life measured using the International Hip Outcome Tool 33 (iHOT-33) (0-100, where a higher score represents a better quality of life) from baseline to 6 months. The iHOT-33 has acceptable psychometric properties and is recommended for use in active adults with hip-related pain.[2 5] The iHOT-33 has a reported minimal clinically important difference of 6.1 points.[2] The minimal important change has been reported as 8.7 and 10.0 points.[6 7]

Secondary outcomes

Unilateral one-repetition-maximum (1 RM) leg press

Unilateral one-repetition-maximum (1 RM) leg press is used as a composite outcome measure for lower limb muscle strength and is measured unilaterally by a 1 RM test in a leg press resistance training machine.[8] Results will be in kilograms.

One-Leg Rise Test (OLRT)

The OLRT is a reliable global measure of lower limb strength and endurance.[9-12] The OLRT is significantly correlated with all strength measures of the hip (i.e., hip flexion, extension, abduction, adduction, internal rotation, and external rotation) in patients with FAIS. Results will be expressed as the number of repetitions, ranging from 0 to 50.

Single leg-hop for distance

The single-leg hop for distance test is a functional performance test evaluating power generation and absorption, expressed in centimeters.[13]

Adherence and exercise volume

Adherence to the exercise sessions (i.e., session adherence reported as the number of attended sessions expressed as a % of the total number of prescribed supervised and total sessions[14]) and exercise progression, sets, repetitions, load, rate of perceived effort (RPE), duration of the exercise session, and whether the session was conducted under supervision, will be registered by the patients as they will be instructed to keep a paper-based exercise

diary.[14] Intensity will be reported as Rating of Perceived Effort (RPE)[15 16]. Volume will be expressed as the product of the total number of repetitions performed and multiplied by the external load.

Patient Acceptable Symptom State (PASS)

PASS is defined as the value beyond which patients consider themselves well, and will be assessed using a single question.

Global Perceived Effect (GPE)

GPE will be assessed for three domains; quality of life, pain, and activities of daily living on a 7-point Likert scale ranging perceived change from 'Worse, an important worsening' (worst) to 'Better, an important improvement' (best).[17 18] The GPE scale asks the patient to rate how much their condition has improved or deteriorated since baseline.

Dropouts:

In instances where patients are not satisfied with how their treatment is progressing before reaching their individual goal, they will be able to have an additional consultation with their treating surgeon, where they will be treated in their best interests. There is a risk of dropout in both groups if the treatment results in insufficient improvement of symptoms. Dropout and the reason for dropout will be recorded throughout the study.

Hip surgery

The number of patients undergoing hip surgery in each intervention group will be recorded throughout the study and registered at 3, 6, and 12 months.

Other outcomes

International Hip Outcome Tool 33 (iHOT-33) subscales:

Each of iHOT-33's subscales (0-100): Symptoms and Functional Limitations, Sports and Recreational activities, Job-related concerns, and Social, Emotional, and Lifestyle concerns has been found valid in patients with hip/groin pain who were not seeking surgery and will be reported.[6]

Copenhagen Hip and Groin Outcome Score (HAGOS)

HAGOS subscales: Pain, Physical function in Sport and Recreation, and QoL (0-100, 0 equals extreme hip and groin problems and 100 equals no hip or groin problems) will be reported. HAGOS has been found valid, reliable, and responsive to assess hip and groin function in physically active adult patients.[19]

The Hip Sports Activity Scale (HSAS)

The Hip Sports Activity Scale (HSAS) is a reliable and valid tool to determine sports levels in patients suffering from FAIS.[20] Results will be expressed from 0 (i.e., no recreational or competitive sports) to 8 (i.e., competitive sports at elite level).

Return to Sport (RTS) questionnaire

Furthermore, a consensus statement on return to sport (RTS) stated that RTS should be reported in categories from no return to participation in sport through return to sport and, finally, return to performance.[21]

The short questionnaire to assess health-enhancing physical activity (SQUASH)

SQUASH is used to measure physical activity.[22 23] Based on the reported effort, the intensity of the activities is assigned an intensity score. The reported sports activities are given a MET value based on Ainsworth´s compendium of Physical Activities, and the intensity score is adjusted according to reported effort.[24] Activities are divided into three intensity categories: (1, light) 2 to < 4 MET, (2, moderate) 4 to < 6.5 MET, and (3, vigorous) ≥ 6.5 MET, and the time per week spent in each category is calculated. The activity score for each item is calculated by multiplying the intensity score and time per week. The total activity score is calculated as the sum of activity scores. Activities with intensity lower than 2 MET will not be included as they are considered to contribute negligibly to physical activity level.[22 23]

Tampa Scale of Kinesiophobia 17 items (TSK-17)

The Tampa Scale of Kinesiophobia (TSK) is a valid and reliable tool to assess somatic focus and activity avoidance in patients. TSK-17 consists of 17 statements that measure pain-related fear of movement in patients with chronic musculoskeletal pain.[25] Each

statement is provided with a 4-point Likert scale, and total scores range from 17 to 68, with a higher score indicating more fear of movement.

Health Economic Evaluation:

We will conduct a cost-effectiveness and cost-utility analysis comparing the physiotherapist-led strength exercise intervention to usual care and estimating the incremental costs per unit of health outcome gained. The health outcomes will be health-related QoL measured using the EQ-5D-5L and iHOT-33.[26 27] Costs will be measured using the Health care utilization questionnaire (HUQ)[28 29] and the Productivity Costs Questionnaire (iPCQ)[30].

EuroQol 5-dimension 5-Level (EQ-5D-5L) and EQ-VAS

Health-related QoL will be assessed using the reliable and valid EuroQol Group 5-dimension 5 Level instrument (EQ-5D-5L), including the summary index ranging from −0.624 (worst) to 1 (best) (Danish value set).[26 27] The EQ-5D-5L is a generic instrument for describing and valuing health. It is based on a descriptive system that defines health in terms of 5 dimensions: Mobility, Self-care, Usual activities, Pain/discomfort, and Anxiety/depression. Each dimension has 5 response categories corresponding to no problems, slight problems, moderate problems, severe problems, and extreme problems. The instrument is designed for self-completion, and respondents also rate their overall health on a vertical visual analog scale (EQ-VAS) ranging from 0 (worst imaginable health) to 100 (best imaginable health). The EQ-5D-5L and EQ-VAS have been widely tested and used in both the general population and patient samples.[27]

Healthcare utilization questionnaire (HUQ):

Healthcare utilization and medication usage are measured by a nine-item patient-reported cost questionnaire (HUQ) including visits to primary care physician and therapist, medical specialist, outpatient hospital care, hospital admissions, diagnostic tests, and medical procedures.[28 29]

The Productivity Costs Questionnaire (iPCQ):

The Productivity Cost Questionnaire (iPCQ) includes three modules measuring productivity losses due to 1) absenteeism, 2) presenteeism of paid work, and 3)

productivity losses related to unpaid work.[30] The iPCQ consists of 18 questions. Productivity losses are measured in three separate modules, allowing the option of leaving out specific types of productivity losses when these are not relevant for a specific patient.

Analysis methods

Primary analysis

Intention-to-treat analysis:

The primary analysis of clinical effectiveness will be the assessment of the between-group mean difference in change in hip-related QoL (iHOT-33) from baseline to 6 months after randomization. Between-group mean differences in change larger than 6.1 points will be considered clinically relevant.[2] Consequently, to determine the intervention's effectiveness compared to usual care, the lower limit of the 95% confidence interval for the mean difference must be greater than 6.1 points. Secondary analysis will be the assessment of the between-group mean difference in change in secondary outcomes from baseline to 3 months, 6 months, and 12 months. Assessment of between-group differences in changes will be conducted using a linear mixed-effect model for continuous variables and a generalized linear mixed-effect model for binary variables. This approach accounts for the correlation between repeated measurements within the same patient over time. The model will include fixed effects for the intervention group (physiotherapist-led exercise or usual care) and time points (Baseline, 3, 6, and 12 months). Patients and recruitment sites (Aarhus, Aalborg, Hvidovre, Horsens, or Melbourne) will be included as random effects to account for potential clustering effects. To assess whether the effect of the intervention on hip-related QoL (iHOT-33) varies over time, an interaction term between the intervention group and time will be included. Supplementary, to further guide the clinical interpretation of change in iHOT-33, the between-group difference in the proportion of participants achieving the minimal clinically important difference of 6.1 for within-patient score change was analyzed using an a-priori threshold of 20%. A difference of less than 20% was considered no meaningful difference between treatments, and 20% or greater was considered a meaningful difference. A sensitivity analysis for thresholds of 10% and 30% will be conducted. Supplementary to this interpretation, we will conduct calculations for other identified values of minimal important change by Scholes et al.[6], of 8.7 points, and Kemp et al.[7], of 10.0 points.

The primary analysis and subsequent inferences will follow the intention-to-treat principle, which means that all participants will be analyzed in the groups to which they were originally randomized, regardless of their adherence to the assigned treatment.

Model validation will be performed by comparing observed and expected within-subject standard deviations and correlations and by inspecting plots of standardized residuals versus fitted values and QQ plots for the standardized residuals.

Per protocol analysis:
Supplementary to the intention-to-treat analysis a per protocol analysis will be conducted. This analysis will focus on participants who strictly adhere (i.e., > 75 % total number of sessions) to the assigned treatment protocol. It will provide additional insights into the treatment effect among participants who complied with the assigned intervention.[31] Furthermore, a sub-analysis will exclude patients who have received hip surgery within 6 months of follow-up.

Missing data
No imputations will be applied in the primary analysis. Each randomized participant will be included in the intention-to-treat analysis with the data collected for the participant. An attempt to collect data from all randomized participants will be made, regardless of adherence to interventions.

Outliers
The dataset will be visually inspected and examined to identify outliers potentially caused by typographical errors. These outliers will be cross-referenced with other data sources (e.g., printed test schemes) or through discussion with the responsible outcome assessor. If confirmed, these outliers will be corrected to actual values. Alternatively, outliers will be removed from the dataset if there is confidence that they are data errors. A sensitivity analysis will be conducted to assess the impact of outlier handling decisions on the statistical analysis results. If the conclusions change substantially due to the exclusion of potential erroneous measurements (which are not confirmed), both conclusions will be reported; otherwise, only the conclusion from the analysis, including all data, will be presented.

Health Economic Evaluation

Incremental cost-effectiveness (ICER) and cost-utility (ICUR) analyses will be conducted from both a societal and healthcare perspective to estimate the cost-effectiveness and cost-utility of supervised strength exercise compared to usual care. The time horizon will be 12 months. Costs will be expressed in Danish Crowns (DKK) (Price year 2026). The base-case cost-effectiveness analysis will follow the intention-to-treat principle. Findings will be reported following the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guideline.[32]

Measurement of costs:

Costs of the supervised strength exercise intervention will be estimated using micro-costing. We will identify, measure, and value direct costs associated with physiotherapist-led strength exercise. These include the time of the physiotherapist, facility, materials, and travel costs. Unit costs for physiotherapists will be identified using salaries from the Danish Association of Physiotherapists in Denmark and the Australian Physiotherapist Association in Australia. Costs in Australian Dollars (AUD) will be converted to DKK using the average purchasing power parity (PPP) exchange rate over the project period (i.e., 2023-2025). Measurement of the number of physiotherapist-led strength exercise sessions will be done using the paper-based exercise diary completed by participants and validated by the physiotherapist delivering the intervention. The value of total costs will be the number of supervised sessions multiplied by the associated salary unit cost. Costs of usual care and other costs in the supervised strength exercise group will be measured using the HUQ questionnaire (e.g., costs related to healthcare and medication usage) and the iPCQ questionnaire (e.g., costs related to absenteeism and presenteeism of paid work, and productivity losses related to unpaid work). Costs will be summarized in Table 4.

Measurement of outcomes:

In the cost-utility analysis, we will use Quality Adjusted Life Years (QALYs) to measure the impact of the healthcare intervention on participants' health-related QoL. QALYs provide a comprehensive and standardized metric that combines both the quantity and quality of life experienced by individuals. To calculate QALYs, we will employ the EQ-5D-5L questionnaire. Using a preference-based algorithm, we will convert EQ-5D-5L responses into a single utility index score, representing each participant's health-related quality of life.

QALYs will be calculated based on the utility index scores over the study period using Danish population norms. We will use linear interpolation between measurement points (baseline, 3 months, 6 months, and 12 months). In the cost-effectiveness analysis, iHOT-33 (0-100) total scores will be used. Outcomes will be summarized in Table 4.

Base-case cost-effectiveness and cost-utility analysis:
Missing data will be imputed using Multiple Imputation with Chained Equations (MICE). Cost and effect data will be assumed to be missing at random, which means that missing observations are explained by observed variables. The imputation model will include outcome variables and predictor variables that differ at baseline, are related to missing data, or are associated with the outcome. To account for the skewed distribution of cost data, predictive mean matching will be used in MICE. The number of imputed datasets will be increased until the loss of efficiency is less than 5%. Each of the imputed datasets will be analyzed separately. Results from the multiple datasets will be pooled using Rubin's rules.[33] To account for the possible clustering of data, analyses will be performed using linear mixed-effect models.[34] Accounting for the possible clustering of data at the hospital level is very important, as most economic evaluations fail to do so, whereas ignoring the possible clustering of data might lead to inaccurate levels of uncertainty and inaccurate point estimates.[34]

The incremental cost-effectiveness ratio (ICER) and cost-utility ratio (ICUR) will be calculated by dividing the between-group difference in change in mean total costs by the difference in mean change for QALYs and iHOT-33 scores. The ICER and ICUR provide a measure of the additional cost per additional QALY, or iHOT-33 point gained or lost by physiotherapist-led exercise compared to usual care. The ICER and ICUR will be presented in Table 5. Bootstrapping techniques will be used to estimate the uncertainty surrounding the cost-effectiveness estimates. Cost-effectiveness planes (Figure 2) visually represent the differences in costs and health outcomes, and cost-effectiveness acceptability curves (Figure 3), graphically present the impact of uncertainty of the ICER and ICUR (the probability of the physiotherapist-led exercise intervention being cost-effective compared to usual care, with possible values of the cost-effectiveness threshold).[35]

Sensitivity analysis:
Various one-way sensitivity analyses will be performed to test the robustness of the study results. First, a complete case analysis will be performed by using only cases with complete data for both outcomes and costs. 20 of the 120 patients enrolled in the study will be from Melbourne, Australia. A sensitivity analysis will be conducted, excluding the patients from Australia because of sample size disparity (100 in Denmark and 20 in Australia) and possible variations in healthcare practices, costs, and outcomes across countries.

Pre-specified exploratory analyses: Adherence and volume

A pre-specified exploratory secondary analysis will investigate whether exercise session adherence (expressed as a % of supervised sessions and as a % of total prescribed sessions) and exercise volume are associated with strength exercise effectiveness on hip-related quality of life (e.g., iHOT-33, 0-100), maximal muscle strength (e.g., 1 RM unilateral leg press) and functional performance (e.g., One Leg Jump for Distance and One Leg Rise Test). Multiple linear regression will be applied in all analyses. With a rule of thumb of 10 observations (i.e., participants) per independent variable, and approximately 60 participants in the physiotherapist-led strength exercise group, we can include 6 independent variables (e.g., categories) in the regression model.[31] The independent variables will be potential moderators measured at baseline and at 6-month follow-up that were expected to be associated with changes in the outcome (e.g., dependent variable). The dependent variable in each of the multiple linear regression models will be the change in the iHOT-33 score (0-100), maximal muscle strength (e.g., 1 RM unilateral leg press), and functional performance (e.g., One Leg Jump for Distance and One Leg Rise Test) from baseline to 6-month follow-up. The independent variables will be 1) Supervised exercise adherence (expressed as a % of maximal 12 sessions), 2) Total exercise adherence (expressed as a % of maximal sessions), 3) Volume, 4) Biological sex (Male/Female), and 5) Age at baseline. This analysis will be reported (Table 6) in a secondary paper with a clear reference to the primary paper.

Statistical software

Data analyses will be conducted using STATA software (StataCorp LP, College Station, Texas), Version 19.1.

Illustrations

Tables

Table 1. Baseline Characteristics.

	Physiotherapist-led exercise (n =)	Usual care (n =)	BetterHip Cohort (n = 120)
Biological Sex (Male/Female)			
Male, n (%)			
Female, n (%)			
Gender			
Male, n (%)			
Female, n (%)			
Specification, n (%)			
Age (years), mean (SD)			
Height (cm), mean (SD)			
Weight (kg), mean (SD)			
Body mass index (kg/m^2), mean (SD)			
FAI Morphology			
CAM, n (%)			
Pincer, n (%)			
Mixed, n (%)			
Alpha Angle, Mean (SD)			
Lateral Center Edge Angle, Mean (SD)			
Positive Crossover Sign, n (%)			
Acetabular Index Angle, Mean (SD)			
Index hip			
Left, n (%)			
Right, n (%)			
Bilateral symptoms			
No, n (%)			
Yes, n (%)			
Hip symptom duration			
0-1 years, n (%)			
1-2 years, n (%)			
2-5 years, n (%)			
> 5 years, n (%)			
Previous treatment due to hip symptoms			
Exercise, n (%)			
Physiotherapy, n (%)			
Chiropractor, n (%)			
Osteopathy, n (%)			
Medicine, n (%)			
Intra-articular steroid injection, n (%)			
Previous hip surgery			
Non-index hip, n (%)			
Educational level			
Primary school, n (%)			
Vocational education, n (%)			
High school or similar, n (%)			
Higher education, n (%)			
Employment status			
Student / Education, n (%)			
Working, n (%)			
Not working, n (%)			
Cohabitation status			
Married or cohabitating, n (%)			
Living alone, n (%)			
Smoking status			
Never, n (%)			
Former, n (%)			
Current, n (%)			

Use of analgesics
Paracetamol, n (%)
NSAIDS, n (%)
Morphine or opioids, n (%)
Other, n (%)
Comorbidities
Specification (%), n (%)
Specification (%), n (%)
Physical activity level (SQUASH)
Total activity score (METs)
Light intensity (min/week)
Moderate intensity (min/week)
Vigorous intensity (min/week)
Pain Kinesiophobia
TSK-17 (17-68), mean (SD)

FAI: Femoroacetabular Impingement. NSAID: Nonsteroidal Anti-inflammatory Drugs. MET: Metabolic Equivalent. SQUASH: the short questionnaire to assess health-enhancing physical activity. TSK-17: Tampa Scale of Kinesiophobia. SD: Standard Deviation. 95% CI: 95% Confidence Interval. MET: Metabolic Equivalent.
If data is not normally distributed median and interquartile range (IQR) will be reported instead of the mean and standard deviation (SD).

Table 2: Changes from baseline to 6-month follow-up in primary and secondary outcomes in the intention-to-treat-population.

	Physiotherapist-led exercise (n =)			Usual care (n =)			
	Baseline (95% CI)	**6-mo (95% CI)**	**Change (95% CI)**	**Baseline (95% CI)**	**6-mo (95% CI)**	**Change (95% CI)**	**Difference in change (95% CI)**
Primary							
iHOT-33, (0-100)							
Secondary							
1-RM Leg Press, Kg							
Index hip							
Non-index hip							
Hop for Distance, cm							
Index hip							
Non-index hip							
One-Leg Rise Test, Reps							
Index hip							
Non-index hip							
PASS (% Yes)							
GPE QoL							
GPE Pain							
GPE ADL							
Hip surgery (%)							

6-mo: 6-month follow-up. 95% CI: 95% Confidence Interval. iHOT-33: International Hip and Outcome Tool 33. 1 RM: 1 Repetition Maximum. Reps: Number of repetitions (0-50). PASS: Patient Acceptable Symptom State. GPE: Global Perceived Effect. QoL: Quality of Life.

Table 3: Changes from baseline to 6-month follow-up in the other outcomes in the intention-to-treat-population.

	Physiotherapist-led exercise (n =)			Usual care (n =)			
	Baseline (95% CI)	**6-mo (95% CI)**	**Change (95% CI)**	**Baseline (95% CI)**	**6-mo (95% CI)**	**Change (95% CI)**	**Difference in change (95% CI)**
iHOT-33 subscales, (0-100)							
Symptoms							
Sport							
Job							
Social							
HAGOS subscales, (0-100)							
Pain							
Sport							
Quality of Life							
SQUASH							
Total activity score (MET)							
Light intensity (min/week)							
Moderate intensity (min/week)							
Vigorous intensity (min/week)							
Sport participation							
HSAS, (1-8)							
RTS, (1-4)							

iHOT-33: International Hip and Outcome Tool 33. Symptoms: Symptoms and Functional Limitations. Sports and Recreational Activities. Job: Job-related concerns. Social: Social, Emotional, and Lifestyle Concerns. HAGOS-Sport: Physical Function in Sport and Recreation. 1 RM: 1 repetition maximum. Cm: centimeters. Kg: Kilograms. PASS: Patient Acceptable Symptom State. HAGOS: The Copenhagen Hip and Groin Outcome Score. SQUASH: the short questionnaire to assess health-enhancing physical activity. TSK-17: Tampa Scale of Kinesiophobia. HSAS: the Hip and Sports Activity Scale. RTS: Return to Sport questionnaire. SD: Standard Deviation. 95% CI: 95% Confidence Interval. MET: Metabolic Equivalent.

Table 4: Costs and effects from baseline to 12-month follow-up in the intention-to-treat population.

	Physiotherapist-led exercise (n =)			Usual care (n =)			
	Baseline (95% CI)	**12-mo (95% CI)**	**Change (95% CI)**	**Baseline (95% CI)**	**12-mo (95% CI)**	**Change (95% CI)**	**ICER/ICUR (95% CI)**
Effects							
QALYs							
iHOT-33, 0-100							
Healthcare costs (HUQ)							
Number of visits							
Physiotherapist							
General practitioner							
Allied health professional							
Complementary care provider							
Emergency department							
Medical specialist							
Admitted to the hospital							
Hip surgery							
Other healthcare providers							
Healthcare Costs (HUQ)							
Medicine usage							
Prescription 1							
Prescription 2							
Prescription 3							
Productivity Costs (IPCQ)							
Indirect costs							
Absenteeism							
Presenteeism							
Productivity loss							

QALY: Quality Adjusted Life Years. HUQ: Health care utilization questionnaire. IPCQ: The Productivity Costs Questionnaire. iHOT-33: International Hip and Outcome Tool 33. SD: Standard Deviation. 95% CI: 95% Confidence Interval. DKK: Danish Crowns.

Table 5. Incremental Cost-Effectiveness Ratios (ICERs), Incremental Cost-Utility Ratios (ICURs), and Distribution of the Joint Cost-Effect Pairs in the Cost-Effectiveness (CE) Plane.

Analysis	Mean cost difference (95% CI)	Mean effect difference (95% CI)	ICUR/ICER	Distribution in CE Plane (%)			
				NE	SE	SW	NW
Main analysis							
iHOT-33							
QALYs							
SA 1: Complete case analysis							
iHOT-33							
QALYs							
SA 2: Imputed data sets							
iHOT-33							
QALYs							
SA 3: Excluding Melbourne							
iHOT-33							
QALYs							

ΔC = mean difference in total costs of the intervention versus usual care. ΔE mean difference in the outcome (iHOT-33 and QALYs) of the intervention versus usual care. ICER = Incremental Cost-Effectiveness Ratio is calculated as $\Delta C/\Delta E$(iHOT-33). ICUR = Incremental Cost-Utility Ratio is calculated as $\Delta C/\Delta E$(QALY). DKK = Danish Kroner. SA = Sensitivity Analysis. QALY = Quality Adjusted Life Years. iHOT-33: International Hip and Outcome Tool 33.
NE: Northeast quadrant of the Cost-Effectiveness Plane: the intervention is more effective and more costly than usual care.
SE: Southeast quadrant of the Cost-Effectiveness Plane: the intervention is more effective and less costly than usual care.
SW: Southwest quadrant of the Cost-Effectiveness Plane: the intervention is less effective and more costly than usual care.
NW: Northwest quadrant of the Cost-Effectiveness Plane: the intervention is less effective and less costly than usual care.
95% CI: 95% Confidence interval

Table 6. Associations between changes in hip-related quality of life (iHOT-33), maximal unilateral muscle strength, functional performance, and potential predictors of training effectiveness (i.e., adherence, exercise volume, biological sex, age, and physical activity).

	Crude analysis,[a, b] β (95% CI)	Adjusted analysis,[a, b, c] β (95% CI)
Model 1: ∆iHOT-33[d]		
$Adherence_{Supervised}$		
$Adherence_{Total}$		
Exercise Volume		
Biological Sex		
Age		
Physical Activity Level		
Model 2: ∆Maximal Unilateral Muscle Strength[d]		
$Adherence_{Supervised}$		
$Adherence_{Total}$		
Exercise Volume		
Biological Sex		
Age		
Physical Activity Level		
Model 3: ∆One Leg Rise Test[d]		
$Adherence_{Supervised}$		
$Adherence_{Total}$		
Exercise Volume		
Biological Sex		
Age		
Physical Activity Level		
Model 4: ∆Single Leg Jump for Distance[d]		
$Adherence_{Supervised}$		
$Adherence_{Total}$		
Exercise Volume		
Biological Sex		
Age		
Physical Activity Level		

[a]Analyzed using linear regression. [b]β, Regression coefficient. The interpretation of the crude and adjusted β is the change in the dependent variable for every unit of each of the independent variables. [c]Adjusted each of the independent variables (i.e., $Adherence_{supervised}$, $Adherence_{Total}$, Exercise Volume, Biological Sex, Age, and Physical Activity Level). [d]Dependent variable in each respective regression model. [95% CI]: 95% confidence interval. (p): P-value. iHOT-33: International Hip and Outcome Tool 33.

Figures

Figure 1. CONSORT Flowchart in The Better Hip Trial.

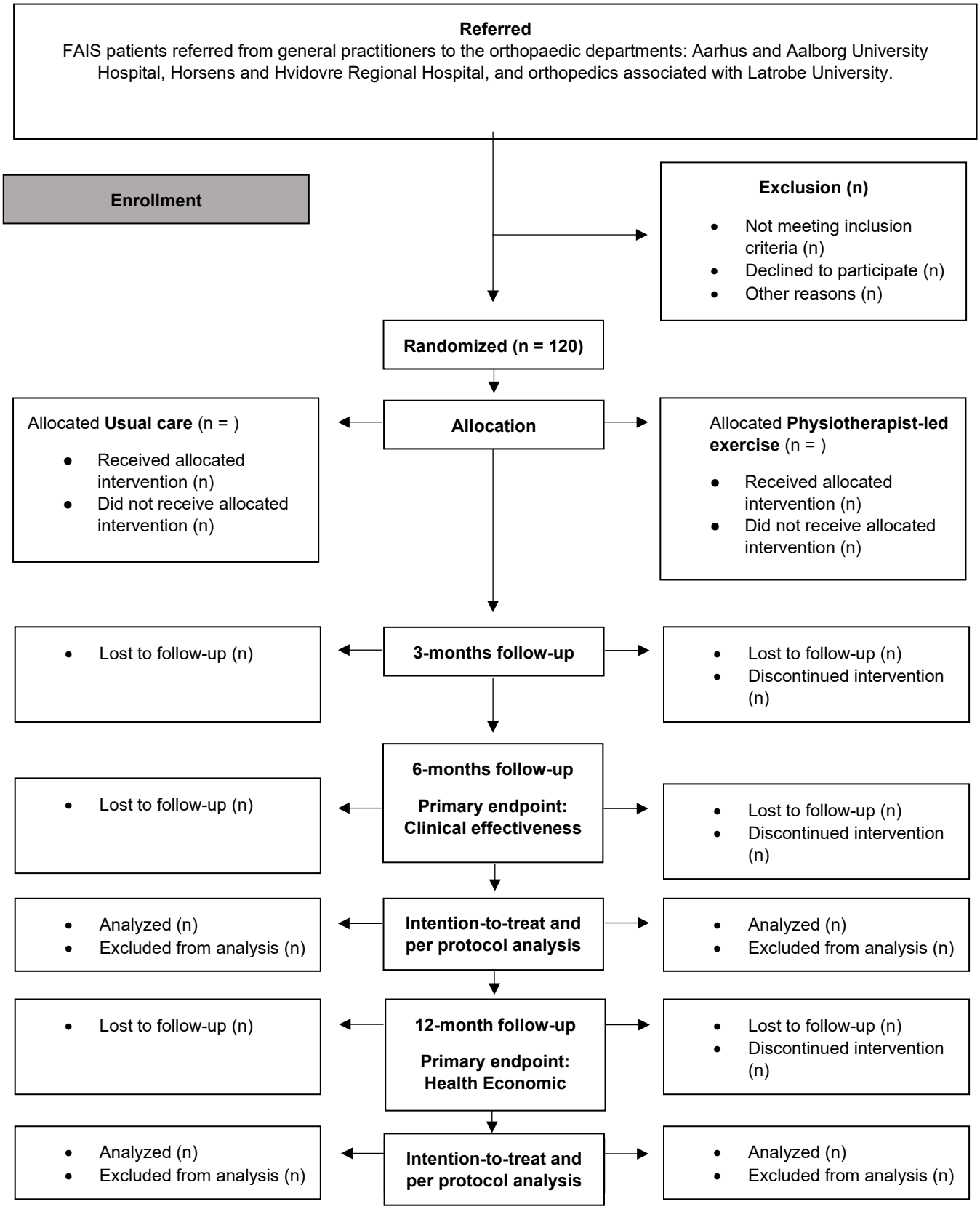


Figure 2. Visualization of changes in iHOT-33 score in the physiotherapist-led exercise group and the usual care group at baseline, 3- and 6-month follow-up with 95% Confidence Intervals. (Example, not based on data).

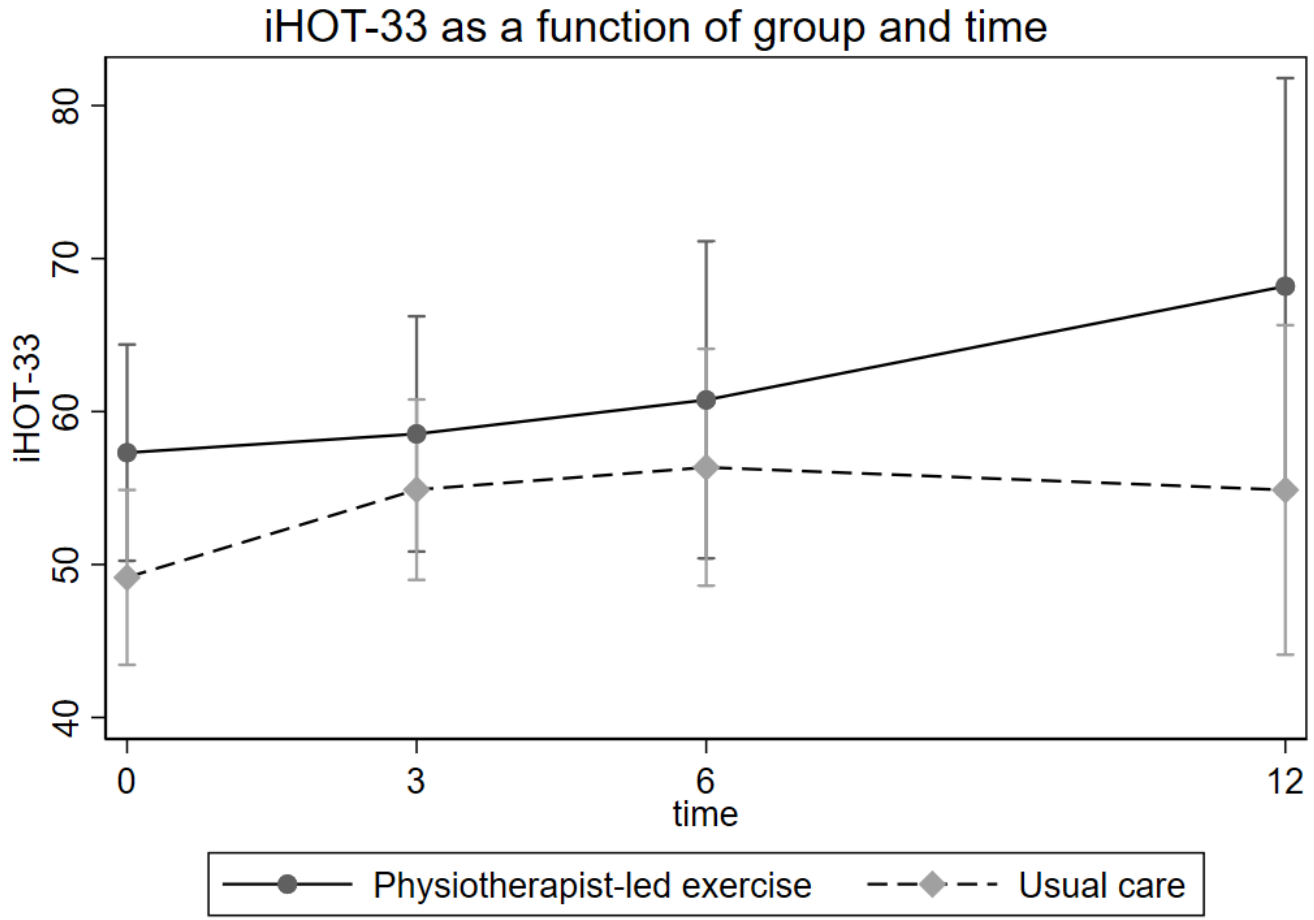


Figure 3. Cost-effectiveness (CE) plane for iHOT-33 and QALYs.

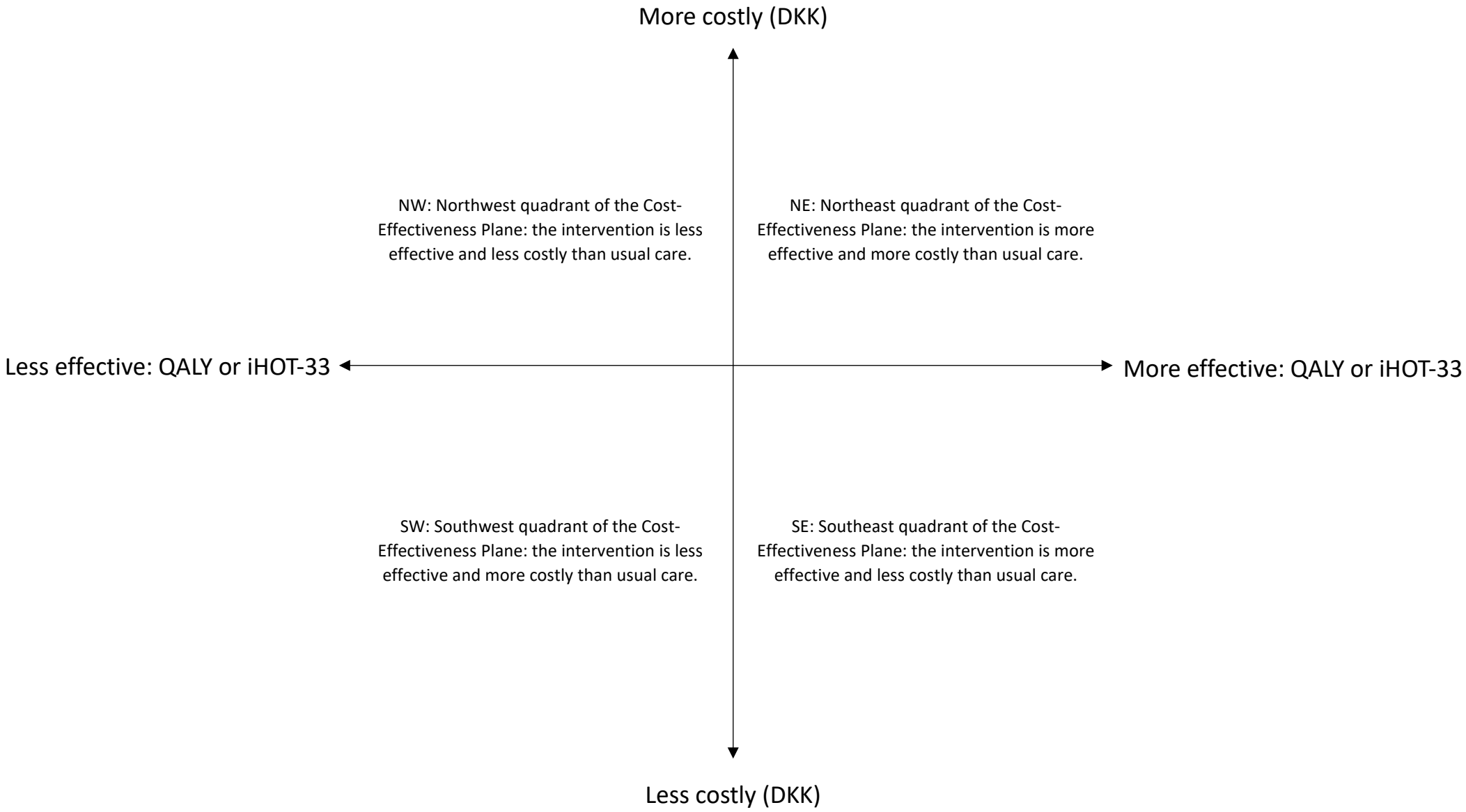
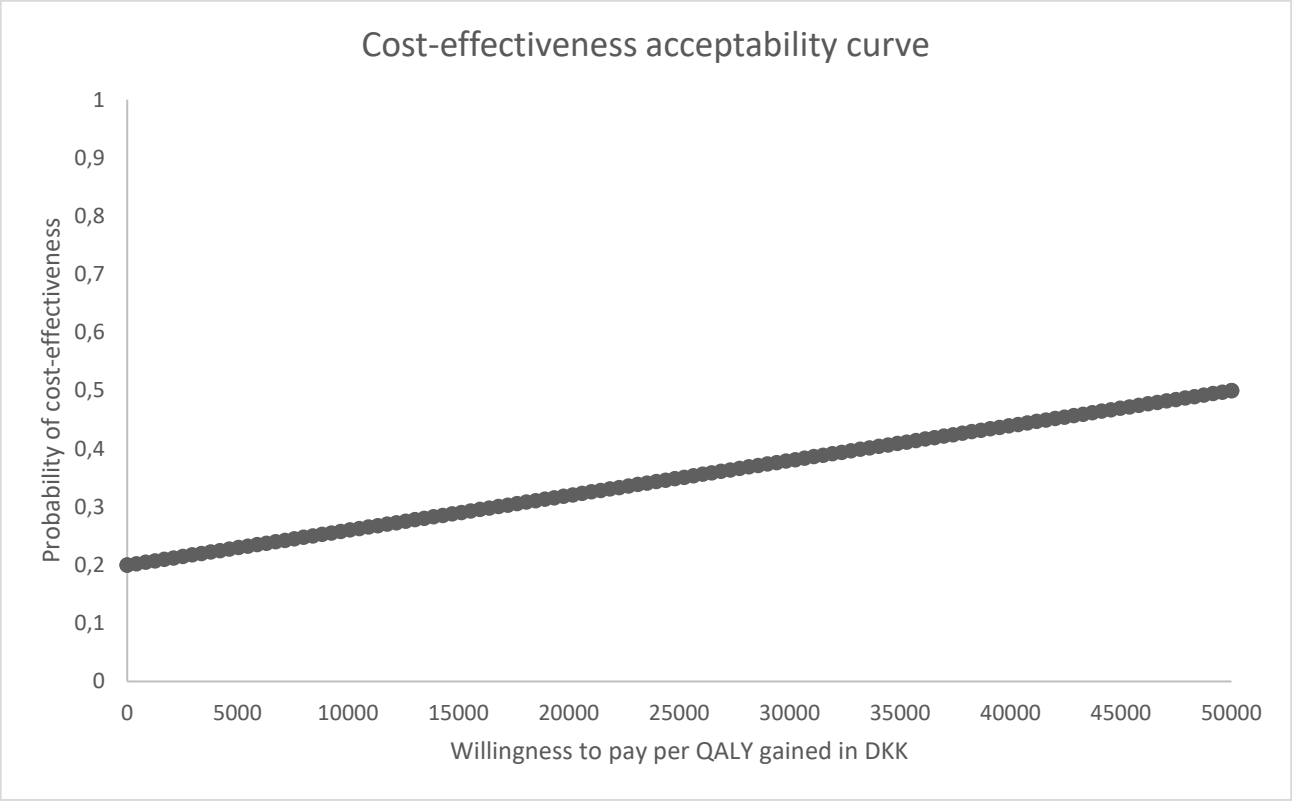


Figure 4. Cost-effectiveness acceptability curve for iHOT-33 and QALYs change (Example, not based on data).



REFERENCES

1. Kemp JL, Coburn SL, Jones DM, et al. The Physiotherapy for Femoroacetabular Impingement Rehabilitation STudy (physioFIRST): A Pilot Randomized Controlled Trial. *J Orthop Sports Phys Ther* 2018;48(4):307-15. doi: 10.2519/jospt.2018.7941
2. Mohtadi NGH, Griffin DR, Pedersen ME, et al. The Development and Validation of a Self-Administered Quality-of-Life Outcome Measure for Young, Active Patients With Symptomatic Hip Disease: The International Hip Outcome Tool (iHOT-33). *Arthroscopy* 2012;28(5):595-610.e1. doi: https://doi.org/10.1016/j.arthro.2012.03.013
3. Foldager FN, Kierkegaard-Brøchner S, Kemp JL, et al. First-line treatment for femoroacetabular impingement syndrome and hip-related quality of life: study protocol for a multicentre randomised controlled trial comparing a 6-month supervised strength exercise intervention to usual care (the Better Hip Trial). *BMJ Open* 2024;14(6):e078726. doi: 10.1136/bmjopen-2023-078726 [published Online First: 20240621]
4. Kemp JL, Johnston RTR, Coburn SL, et al. Physiotherapist-led treatment for femoroacetabular impingement syndrome (the PhysioFIRST study): a protocol for a participant and assessor-blinded randomised controlled trial. *BMJ open* 2021;11(4):e041742-e42. doi: 10.1136/bmjopen-2020-041742
5. Impellizzeri FM, Jones DM, Griffin D, et al. Patient-reported outcome measures for hip-related pain: a review of the available evidence and a consensus statement from the International Hip-related Pain Research Network, Zurich 2018. *Br J Sports Med* 2020;54(14):848-57. doi: 10.1136/bjsports-2019-101456 [published Online First: 20200217]
6. Scholes MJ, King MG, Crossley KM, et al. The Validity, Reliability, and Responsiveness of the International Hip Outcome Tool-33 (iHOT-33) in Patients With Hip and Groin Pain Treated Without Surgery. *Am J Sports Med* 2021;49(10):2677-88. doi: 10.1177/03635465211027180 [published Online First: 20210715]
7. Kemp JL, Collins NJ, Roos EM, et al. Psychometric Properties of Patient-Reported Outcome Measures for Hip Arthroscopic Surgery. *Am J Sports Med* 2013;41(9):2065-73. doi: 10.1177/0363546513494173
8. Thompson WR GN, Pescatello LS. ACSM's guide-lines for exercise testing and prescription. 8 ed: Wolters Kluwer Health 2010.
9. Culvenor AG, Collins NJ, Vicenzino B, et al. Predictors and effects of patellofemoral pain following hamstring-tendon ACL reconstruction. *J Sci Med Sport* 2016;19(7):518-23. doi: 10.1016/j.jsams.2015.07.008
10. Thongchoomsin S, Bovonsunthonchai S, Joseph L, et al. Clinimetric properties of the one-leg sit-to-stand test in examining unilateral lower limb muscle strength among young adults. *Int J Clin Pract* 2020;74(9):e13556. doi: https://doi.org/10.1111/ijcp.13556
11. Kemp JL, Risberg MA, Schache AG, et al. Patients With Chondrolabral Pathology Have Bilateral Functional Impairments 12 to 24 Months After Unilateral Hip Arthroscopy: A Cross-sectional Study. *J Orthop Sports Phys Ther* 2016;46(11):947-56. doi: 10.2519/jospt.2016.6577
12. Kemp JL, Makdissi M, Schache AG, et al. Is quality of life following hip arthroscopy in patients with chondrolabral pathology associated with impairments in hip strength or range of motion? *Knee Surg Sports Traumatol Arthrosc* 2016;24(12):3955-61. doi: 10.1007/s00167-015-3679-4

13. Roughead EA, King MG, Crossley KM, et al. Football players with long standing hip and groin pain display deficits in functional task performance. *Phys Ther Sport* 2022;55:46-54. doi: https://doi.org/10.1016/j.ptsp.2022.02.023
14. Dennett R, Madsen LT, Connolly L, et al. Adherence and drop-out in randomized controlled trials of exercise interventions in people with multiple sclerosis: A systematic review and meta-analyses. *Mult Scler Relat Disord* 2020;43:102169. doi: https://doi.org/10.1016/j.msard.2020.102169
15. Progression Models in Resistance Training for Healthy Adults. *Med Sci Sports Exerc* 2009;41(3)
16. Helms ER, Storey A, Cross MR, et al. RPE and Velocity Relationships for the Back Squat, Bench Press, and Deadlift in Powerlifters. *J Strength Cond Res* 2017;31(2):292-97. doi: 10.1519/jsc.0000000000001517
17. Kamper SJ, Ostelo RWJG, Knol DL, et al. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63(7):760-66.e1. doi: 10.1016/j.jclinepi.2009.09.009
18. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10(4):407-15. doi: https://doi.org/10.1016/0197-2456(89)90005-6
19. Thorborg K, Hölmich P, Christensen R, et al. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med* 2011;45(6):478-91. doi: 10.1136/bjsm.2010.080937
20. Naal FD, Miozzari HH, Kelly BT, et al. The Hip Sports Activity Scale (HSAS) for patients with femoroacetabular impingement. *Hip Int* 2013;23(2):204-11. doi: 10.5301/hipint.5000006 [published Online First: 20130328]
21. Ardern CL, Glasgow P, Schneiders A, et al. 2016 Consensus statement on return to sport from the First World Congress in Sports Physical Therapy, Bern. *Br J Sports Med* 2016;50(14):853-64. doi: 10.1136/bjsports-2016-096278
22. Sørensen L, Mikkelsen LR, Jacobsen JS, et al. Reliability of the Danish version of the short questionnaire to assess health-enhancing physical activity (SQUASH). *Physiother Theory Pract* 2018;34(8):1-642. doi: 10.1080/09593985.2017.1423143
23. Wendel-Vos GCW, Schuit AJ, Saris WHM, et al. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol* 2003;56(12):1163-69. doi: 10.1016/S0895-4356(03)00220-8
24. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;43(8):1575-81. doi: 10.1249/MSS.0b013e31821ece12
25. Miller RP, Kori SH, Todd DD. The Tampa Scale: a Measure of Kinisophobia. *Clin J Pain* 1991;7(1):51. doi: 10.1097/00002508-199103000-00053
26. Wittrup-Jensen KU, Lauridsen J, Gudex C, et al. Generation of a Danish TTO value set for EQ-5D health states. *Scand J* 2009;37(5):459-66. doi: 10.1177/1403494809105287
27. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20(10):1727-36. doi: 10.1007/s11136-011-9903-x
28. Goossens ME, Rutten-van Molken MP, Vlaeyen JW, et al. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*

2000;53(7):688-95. doi: 10.1016/s0895-4356(99)00177-8 [published Online First: 2000/08/15]
29. van den Brink M, van den Hout WB, Stiggelbout AM, et al. Self-reports of health-care utilization: Diary or questionnaire? *Int J Technol Assess Health Care* 2005;21(3):298-304. doi: 10.1017/S0266462305050397
30. Bouwmans C, Krol M, Severens H, et al. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. *Value Health* 2015;18(6):753-8. doi: 10.1016/j.jval.2015.05.009 [published Online First: 2015/09/28]
31. Kirkwood BR, Sterne JAC. Essential medical statistics. 2. ed. ed. Malden, Mass: Blackwell Science 2012.
32. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *BMJ* 2022;376:e067975. doi: 10.1136/bmj-2021-067975
33. White IR RP, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med 2011; 30: 377–399* 2011
34. Gomes M, Grieve, R., Nixon, R., Ng, E. S., Carpenter, J., & Thompson, S. G. Methods for covariate adjustments in cost-effectiveness analysis that use cluster randomised trials. . *Health Economics, 21(9), 1101– 1118* 2012
35. Boden I, Elkins MR. Economic evaluations of physiotherapy interventions. *J Physiother* 2023;69(3):136-38. doi: https://doi.org/10.1016/j.jphys.2023.05.004