Total Hip Arthroplasty *versus* Progressive Resistance Training in Patients with Severe Hip Osteoarthritis:

Statistical Analysis Plan for the PROHIP multicentre, parallel-group, randomised trial

Thomas Frydendal[1,2], Robin Christensen[3,4], Inger Mechlenburg[5,6], Lone Ramer Mikkelsen[6,7], Søren Overgaard[8,9], Kim Gordon Ingwersen[1,10]

Affiliations:

[1]Department of Physiotherapy, Lillebaelt Hospital – University Hospital of Southern Denmark, Vejle Hospital, Denmark.
[2]Department of Clinical Research, University of Southern Denmark, Odense, Denmark.
[3]Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
[4]Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense Denmark
[5]Department of Orthopaedic Surgery, Aarhus University Hospital, Aarhus, Denmark
[6]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[7]Elective Surgery Centre, Silkeborg Regional Hospital, Silkeborg, Denmark
[8]Department of Orthopaedic Surgery and Traumatology, Copenhagen University Hospital, Bispebjerg Hospital, Copenhagen, Denmark
[9]Department of Clinical Medicine, Faculty of Health and Medical Sciences University of Copenhagen, Copenhagen, Denmark
[10]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark.

Principle Investigator: Thomas Frydendal (TF), Department of Physiotherapy, Lillebaelt Hospital – University Hospital of Southern Denmark, Vejle Hospital, Denmark and Department of Clinical Research, University of Southern Denmark, Odense, Denmark.

Senior Biostatistician and Analysis Support: Robin Christensen (RC), Professor of Biostatistics & Clinical Epidemiology: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark and Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense Denmark

Statistical Analyst: Thomas Frydendal (TF), Department of Physiotherapy, Lillebaelt Hospital – University Hospital of Southern Denmark, Vejle Hospital, Denmark and Department of Clinical Research, University of Southern Denmark, Odense, Denmark.

Database Manager: David Hass (DH), Open Patient data Explorative Network (OPEN), Odense University Hospital, Region of Southern Denmark

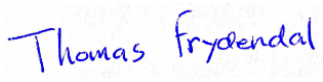
Trial Registration Number: ClinicalTrials.gov (NCT04070027)

Protocol version: This document has been written based on the information contained in the trial protocol version 1.3, dated 22th of June 2021

STATISTICAL ANALYSIS PLAN (SAP) REVISION HISTORY

Protocol version	SAP version	Section changed	Description of and reason for change	Date changed
1.3	1.0			26-05-2021

SIGNATURES

Date	Name	Roles and Responsibility	Signature
25-06-2021	Thomas Frydendal	SAP first author	
25-06-2021	Robin Christensen	Senior Biostatistician	
25-06-2021	Søren Overgaard	Chief Investigator	

INTRODUCTION

Background and rationale

In short, hip osteoarthritis (OA) is the leading cause for total hip arthroplasty (THA) with more than one million procedures performed annually worldwide.[1] Although, the procedure is considered as the surgery of the century up to 23% of the patients report long-term pain and deficits in physical function and muscle strength may persist after THA.[2-6]

Exercise is recommended as first-line treatment,[7-11] and progressive resistance training (PRT) appears to moderately improve multiple outcomes, and may be of clinical relevance in patients with hip OA.[12-15] Current treatment selection in patients with hip OA is based on low-level evidence as no randomised controlled trials have directly compared THA to non-surgical treatment.[9] This comparison is important in order to ensure an evidence based treatment and improve shared decision making for patients with hip OA.

Objectives

The primary objective of this trial is to determine the effectiveness of THA followed by standard care, compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on change in hip pain and function, measured using the Oxford Hip Score (OHS) assessed after 6 months, in patients with severe hip OA.

The primary null hypothesis is that there is no difference in hip pain and function between the two groups. The primary alternative hypothesis is that patients randomised to THA followed by standard care will improve significantly and clinically relevant more in hip pain and function 6 months after initiating the treatment than those randomised to PRT.

Secondary objectives are:

1) To determine the effectiveness of THA followed by standard care compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on changes in hip pain, hip symptoms, activities-of-daily-living (ADL) function, hip-related quality-of-life (QoL), and sport and recreation function (sport/recreation), measured using the Hip disability and Osteoarthritis Outcome Score (HOOS) subscales, from baseline to the 6-months follow-up.

2) To determine the effectiveness of THA followed by standard care compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on change in physical activity level, measured using University of California Los Angeles (UCLA) Activity Score, from baseline to the 6-months follow-up.
3) To determine the effectiveness of THA followed by standard care compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on change in gait speed, measured using The 40m fast-paced walk test (40m-FPWT), from baseline to the 6-months follow-up.
4) To determine the effectiveness of THA followed by standard care compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on change in sit-to-stand function, measured using the 30-second chair stand test (30s-CST), from baseline to the 6-months follow-up.
5) To determine the safety of THA followed by standard care compared to 12 weeks of supervised PRT followed by 12 weeks of optional unsupervised PRT, on number of serious adverse events (SAEs) from baseline to the 6-months follow-up.

STUDY METHODS

Trial design

This trial is a multicentre (four sites), stratified (by site), randomised (allocation 1:1), controlled, parallel-group superiority trial. Patients are recruited from the orthopaedic departments at Vejle Hospital and Odense University Hospital (OUH) in the Region of Southern Denmark, Aarhus University Hospital (AUH) in the Central Denmark Region, and Næstved Hospital in Region Zealand and randomised to THA followed by standard care or 12-weeks of supervised PRT followed by 12-weeks of optional unsupervised PRT. The treatments are described in full details in the trial protocol. The primary endpoint is the 6-months follow-up, while a secondary endpoint is the 3-months follow-up.

The statistical analysis plan (SAP) and trial protocol are reported in accordance with the *"Guidelines for the Content of Statistical Analysis Plans in Clinical Trials"*[16] and *"Standard Protocol Items: Recommendations for Interventional Trials"* (SPIRIT), respectively,[17] while reporting of the trial will follow the *"Consolidated Standards of Reporting Trials"* (CONSORT) statement.[18]

Trial registration was performed at ClinicalTrials.gov (NCT04070027) in August 2019. Patient enrolment started at the first hospital in September 2019 and at the last hospital in February 2020. The trial has been approved by The Regional Committees on Health Research Ethics for Southern Denmark (Project-ID: S-20180158) and the Danish Data Protection Agency (Journal No 19/20337).

Randomisation

Patients are randomised after baseline assessment to THA or PRT with a 1:1 allocation as per a computer-generated randomisation schedule, stratified by recruitment site using permuted blocks of varying sizes (two to six). An independent data manager developed the computer-generated list of random numbers using the randomisation tool in Research Electronic Data Capture (REDCap).[19] Administrators of the randomisation procedure is blinded to block sizes and randomisation sequence at all times during the trial period. The randomisation code is stored in REDCap with no access from the author group.

Sample size and power calculation

The sample size and power estimation was based on the expected between-group difference in the OHS mean change score from baseline to the 6-months follow-up. The OHS mean baseline value is anticipated to be between 14 and 20 points.[20-23] For the OHS, the minimal clinically important difference in the change score between two groups has been estimated to be 5 points and standard deviation (SD) of the change score to be approximately 8 points from baseline to 6 months after THA.[20] Both groups are expected to experience improvements corresponding to a 20 point mean improvement in the THA group as reported in previous studies,[20 21] and up to a 15 point mean improvement in the PRT group comparable with effects of previous exercise interventions.[13 15]

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 SD change of 8, a sample size of 60 per group has a power of 0.92 for the primary outcome to detect a mean change difference of 5 OHS points after 6 months between the THA and PRT group. To obtain at least 80% power to detect a between-group difference in mean change of 5 OHS points with a SD change of 8 OHS points, a sample size of 42 per group will be required.

Statistical interim analysis and stopping guidance

No formal statistical interim analysis is planned on the primary endpoint (baseline to 6-months follow-up) for the OHS change score between the two groups (THA and PRT). The final deadline for patient recruitment was a priori set 18 months (i.e. February 2021) after the inclusion of patients was started. However, due to COVID-19 lockdown in Denmark in 2020 the recruitment deadline was prolonged to June 2021. The author group have monitored recruitment and attrition rates in the trial.

Timing of final analysis

The final analysis for the between-group comparison (THA vs. PRT) for the primary endpoint (baseline to 6-months follow-up) is planned to be performed after each randomised patient has completed the 6-months follow-up. The main publication of the trial will be prepared when these data have been received and cleaned (anticipated by March 2022).
In subsequent manuscripts, secondary longer-term endpoints will be analysed when the 24-months (anticipated by September 2023) and 60-months follow-up (anticipated by September 2026) have been reached for all randomised patients followed by preparation of manuscripts with two and five-year outcomes, respectively.

Timing of outcome assessments

The overview of trial procedures and time-point of each outcome assessment is presented in the Table 3 in the trial protocol. The intervention start date (THA or PRT) for each patient is used to calculate follow-up time points. Then, 12 and 26 weeks are added to determine the expected dates for the 3 and 6 months follow-up, respectively.

STATISTICAL PRINCIPLES

Confidence intervals and p-values

All statistical tests and confidence intervals will be two-sided and $p\text{-}value < 0.05$ will be considered statistically significant; confidence intervals will be based upon 95% (95% CI). A 95% confidence interval excluding a difference greater than 5 OHS points between the two groups will be interpreted as indicating absence of a minimal clinically important difference (i.e. possible equivalence). To account for multiplicity, the analyses of the key secondary outcomes will be performed in prioritized order (i.e. gatekeeping procedure) until one of the analyses fails to show a

statistically significant difference, or until all analyses have been completed at a statistical significance level of $p < 0.05$ using the following order:

1) HOOS pain
2) HOOS symptoms
3) HOOS ADL function
4) HOOS QoL
5) HOOS sport/recreation
6) UCLA Activity Score
7) 40m-FPWT
8) 30s-CST
9) SAEs.

Adherence and protocol deviations

In the THA group, adherence is based on the number and percentage of patients undergoing the scheduled surgery, with the latter defined as: adherence% = (number of patients undergoing THA / number of planned THA in the THA group) x 100%.

In the PRT group, adherence is based on the number and percentage of PRT sessions completed, with the latter defined as: adherence% = (number of PRT sessions completed / total number of planned PRT sessions according to the trial protocol) x 100%. High adherence is defined as participation in ≥75% of the sessions (i.e. 18 out of 24 sessions); moderate adherence as participation in 50-74% of the sessions; and poor adherence as participation in <50% of the sessions.[24] Treatment adherence will be presented by randomisation group using descriptive statistics (number (percentage)).

The following are pre-defined major protocol deviations: (1) patients in the PRT group undergoing THA between baseline and 6-months follow-up, (2) patients in the THA group declining surgery after randomisation, and (3) patients in both groups withdrawing from the trial between baseline and 6-months follow-up. The number (percentage) of patients with major protocol deviations will be summarised by treatment group. No formal statistical testing will be conducted.

Analysis populations

The primary analyses will be based on the Intention to Treat (ITT) population. The ITT principle asserts the effect of a treatment policy (that is, the planned treatment regimen operationalised via randomisation), rather than the actual treatment given (i.e., it is independent of treatment adherence). Accordingly, participants allocated to a treatment group (THA and PRT, respectively) 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).[25 26]

A per-protocol analysis will be conducted including patients in the THA group undergoing surgery and patients in the PRT with sufficient adherence to the exercise program (i.e. participation in ≥75% of the sessions).

An as-treated analysis will be performed based on the patients' adherence to the randomised treatment expecting up to four groups:

1) Patients randomised to THA undergoing surgery.
2) Patients randomised to PRT without undergoing THA in the follow-up period.
3) Patients randomised to THA but declined surgery post randomisation.
4) Patients randomised to PRT undergoing THA during the follow-up period.

The per-protocol and as treated analyses will be considered as a combination of sensitivity and exploratory analyses with the purpose to evaluate the robustness of the ITT analysis.

TRIAL POPULATION

Screening data

The total number of patients screened for eligibility from the four hospitals will be collected and presented in a CONSORT flowchart to describe representativeness of the trial sample. Furthermore, the number of ineligible patients randomised by mistake, if any, will be reported including reason for ineligibility.

Eligibility

Patients conforming to the following inclusion and exclusion criteria are considered eligible for the trial.

Inclusion criteria: (1) Patients aged ≥50 years; (2) Clinical history and symptoms consistent with primary hip OA (including hip OA due to mild hip dysplasia that may be treated with standard components) and radiographic verified hip OA defined as joint space narrowing <2 mm; (3) Considered eligible for THA by an orthopaedic surgeon (i.e. hip-related pain, symptom duration >3 months, functional impairment or decreased range-of-motion, and attempted treatment with analgesics).

Exclusion criteria: (1) Severe walking deficits (i.e. dependency of two crutches or walker); (2) Body Mass Index (BMI) >35 kg/m^2; (3) Lower extremity fractures within previous 12 months; (4) Planned other lower extremity surgery within 6 months; (5) Cancer diagnosis and current chemo-, immuno- or radiotherapy; (6) Neurological diseases (e.g. previous stroke, multiple sclerosis, Parkinson's, Alzheimer's); (7) Other reasons for exclusion (i.e. inadequacy in written and spoken Danish, mentally unable to participate, physically unable to comply with the PRT protocol due to comorbidity (e.g. severe heart disease, previous major lower extremity surgery within previous 6 months) etc.).

Recruitment

The CONSORT flowchart will comprise number of patients screened, excluded (with reasons) eligible for inclusion in the trial, randomised, receiving their allocated treatment, withdrawals (with reasons), and lost to follow-up (with reasons), included in ITT analysis, included in per protocol analysis, and included in as treated analysis. The CONSORT flowchart is depicted in **Figure 1.**

Withdrawal/follow-up

The level of consent withdrawal will be classified as the following two options: (1) consent to continue follow-up and data collection and (2) complete withdrawal with no further follow-up and data collection.

Timing of withdrawal and loss to follow-up will be presented in the CONSORT flowchart with numbers and reasons for withdrawal and/or loss to follow-up given at the 3-months and 6-months (primary end point) outcome assessment. Furthermore, the number (with reasons) of loss to follow-up during the course of the trial will summarised by treatment group.

Baseline patient characteristics

The following data will be used to describe patients by randomisation group at baseline: sex, age, height, weight, body mass index, educational level beyond high school, employment status, alcohol consumption above Danish recommendations, index hip, duration of hip symptoms, previous THA, previous total knee arthroplasty, medicine consumption in the previous week, OHS, HOOS subscales (pain, symptoms, ADL function, sport/rec, hip-related QoL), UCLA activity score, 40m-FPWT, and 30s-CST.

Numbers and percentages will be calculated and presented for categorical variables. Means and SD will computed and presented for continues variables if data follows a normal distribution. In case, continuous variables are not normally distributed, median and interquartile-range will be calculated. No tests of statistical significance will be conducted for any baseline characteristic variable. However, imbalances with clinical importance will be noted. The baseline characteristics will be presented as illustrated in **Table 1.**

ANALYSIS

Outcome definitions

Primary outcome

Oxford Hip Score (OHS)

The primary outcome measure will be the between-group difference in change from baseline to 6-months follow-up in the OHS ($\Delta OHS_{6\text{-months follow-up}} - OHS_{baseline}$). The OHS is considered a valid, reliable, and responsive patient-reported questionnaire assessing hip pain and function in a composite score ranging from 0 (worst) to 48 (best).[27 20 23 28 29]

Response to treatment will be computed for the OHS change score for each patient in both treatment groups and presented dichotomized (i.e. responder and non-responder) as number (and percentages) responders. The following two methods will be used to assess treatment responders:

1) Patients will be classified as an OHS minimal important clinically responder if the OHS change score improves by 8 points or more (≥) from baseline to 6-months follow-up.[20]
2) According to the criteria of the Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative (OMERACT–OARSI),[30] patients will be classified as a

responder if the OHS pain or function subscale score improves by 50% or more and by 20 or more absolute points; *OR* if two of the following three findings is observed: an improvement in the OHS pain score by 20% or more and by 10 or more absolute points on the subscale, an improvement in the OHS function score by 20% or more and by 10 or more absolute points on the subscale, or an improvement in the score on the patient's OHS score by 20% or more and by 10 or more points on the scale from baseline to 6-months follow-up. The OHS pain and function subscales ranges from 0 to 100, with higher scores indicating better disease status.[31]

Key secondary outcomes

Hip disability and Osteoarthritis Outcome Score (HOOS)

A key secondary outcome will be the between-group difference in change from baseline to 6-months follow-up in each HOOS subscale ($\Delta HOOS_{6\text{-months follow-up}} - HOOS_{baseline}$).The HOOS is a valid, reliable and responsive patient-reported questionnaire consisting of five subscales covering, hip pain, hip symptoms, ADL function, hip-related QoL, and sport/recreation with each subscale score ranging from 0 (worst) to 100 (best).[32-35]

University of California Los Angeles (UCLA) Activity Score

A key secondary outcome will be the between-group difference in change from baseline to 6-months follow-up in UCLA activity score (ΔUCLA activity $score_{6\text{-months follow-up}} -$ UCLA activity $score_{baseline}$). The UCLA is reliable, valid, and responsive measure of patient-reported physical activity level ranging from 1 (inactive) to 10 (regular participation in impact sport or heavy labour).[36-38]

The 40m fast-paced walk test (40m-FPWT)

A key secondary outcome will be the between-group difference in change from baseline to 6-months follow-up in the 40m-FPWT ($\Delta 40\text{m-FPWT}_{6\text{-months follow-up}} - 40\text{m-FPWT}_{baseline}$). The 40m-FWT is considered a valid, reliable and responsive test for assessing short distance maximum walking speed.[39] The unit of the change score will be presented as [meters/seconds].

The 30-second chair stand test (30s-CST)

A key secondary outcome will be the between-group difference in change from baseline to 6-months follow-up in the 30s-CST (Δ30s-CST$_{\text{6-months follow-up}}$ - 30s-CST$_{\text{baseline}}$). The 30s-CST is a valid, reliable, and responsive test for evaluating sit-to-stand function measuring the number of sit-to-stand repetitions completed within 30 seconds.[39-42] The primary and key secondary outcomes will be presented as illustrated in **Table 2.**

Analysis methods

All descriptive statistics and statistical analysis will be reported in accordance with the recommendations of the *"Enhancing the QUAlity and Transparency Of health Research"* (EQUATOR) network[43] and the CONSORT statement.[18] Visual inspection (QQ-plot, histograms, and scatterplots) of the standardised residuals from the statistical model will be used to assess the assumption of normality and homogeneity of variances.

Between-group differences of continuous outcomes will be estimated using repeated-measures mixed effects linear models. Data will be analysed with each outcome variable (Y_i) at baseline ($Y_{0,i}$) as a covariate, using a multi-level repeated measures mixed effects model with patients as the random effects factor based on a restricted maximum likelihood (REML) model. Change from baseline to the 6-months follow-up will be the dependent variable, and baseline value (one for each patient), treatment group (two levels: THA and PRT), and time-point (three levels: baseline, 3, and 6 months), hospital (four levels: Vejle, OUH, AUH, and Næstved) will be included as (fixed effect) covariates, as well as the interaction between treatment group and time. This statistical model include all between-group comparisons at both outcome assessments following baseline, which also allows for evaluation of the average effect (i.e. Group as a main effect), as well as the trajectory over time from baseline to 6-months follow-up (i.e. Group×Time interaction). Categorical outcomes will be analysed with logistic regression using identical fixed effect factors and covariates as the mixed linear model.

Sensitivity and exploratory analyses will be performed with the purpose to test the robustness of the primary analyses, including a per-protocol (i.e. surgery performed in the THA group and participation in ≥75% of the training sessions in the PRT group) and as-treated analysis, in which patients will be analysed based on their adherence to the randomised treatment expecting four groups: (1) patients randomised to THA undergoing surgery, (2) patients randomised to PRT without undergoing THA in the follow-up period, (3) patients randomised to THA but declined

surgery post randomisation, and (4) patients randomised to PRT undergoing THA during the follow-up period.

In addition, subgroup analyses will be performed to examine whether the observed treatment effect varies across patient subgroups on the primary outcome are planned from baseline to 6-months follow-up, to explore whether the overall treatment effect is modified by the value of a variable assessed at baseline: analysed by sex (male/female), median age, obesity (BMI ≥30/BMI ≤29.9), median duration of hip symptoms, previous THA (yes/no), median OHS, median UCLA activity score, median walking speed in the 40m-FPWT, and median sit-to-stand repetitions in the 30s-CST. The statistical approach for this evaluation of potential effect modifiers will be a test for statistical interaction to evaluate whether the treatment effect varies across levels of the effect modifier.[44]

Finally, blinded results from the statistical analyses (Group A compared to Group B) will be presented to the author group followed by development of two written interpretations. The author group will sign a consensus statement comprising both interpretations prior to the unsealing of the randomisation code.[45]

Missing data and sensitivity analyses

Missing data will be handled indirectly by relying on the statistical model based on repeated-measures linear mixed models: These models are valid if data are "*Missing at Random*" (i.e. any systematic difference between the missing values and the observed values can be explained by differences in observed data).[46]

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. The following four point framework[47] for rigorous interpretation of the impact of missing data will be applied in the ITT analysis:

1) Attempt to follow up all randomised patients, even if they withdrew 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 missing data (i.e., Model-based: data as randomised; using 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 single-step non-responder-imputation: using the value at baseline to replace missing data; these models will be informative to assess robustness in case data are "*Missing Not At Random*" [MNAR]).
4) Account for all randomised patients, at least in the sensitivity analyses (covered by #2 and #3 above, plus the corresponding analyses based on the per protocol population).

Additional analyse

No additional analyses on the primary and key secondary outcomes are planned from baseline to 6-months follow-up.

Harms

SAEs will be defined in accordance to the *"International Conference on Harmonisation-Good Clinical Practice"* (ICH-GCP) guidelines.[48] In the PRT group, crossover to THA will not be classified as an SAE. The number (and percentage) of occurrences of all SAEs and discontinuation due to SAEs will be presented for each group. Statistical testing will be conducted using logistic regression as described in the "*Analysis methods"* section. The SAEs will be presented as illustrated in **Table 3.**

Statistical software

All statistical analyses and calculations will be performed using SAS (SAS Institute Inc., Cary, North Carolina, USA) and STATA (Statacorp, College Station, Texas, USA).

REFERENCES

1. Ferguson RJ, Palmer AJ, Taylor A, et al. Hip replacement. *Lancet* 2018;392(10158):1662-71. doi: 10.1016/s0140-6736(18)31777-x [published Online First: 2018/11/30]
2. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370(9597):1508-19. doi: 10.1016/s0140-6736(07)60457-7 [published Online First: 2007/10/30]
3. Jensen C, Aagaard P, Overgaard S. Recovery in mechanical muscle strength following resurfacing vs standard total hip arthroplasty - a randomised clinical trial. *Osteoarthritis Cartilage* 2011;19(9):1108-16. doi: 10.1016/j.joca.2011.06.011 [published Online First: 2011/07/14]
4. Rasch A, Dalen N, Berg HE. Muscle strength, gait, and balance in 20 patients with hip osteoarthritis followed for 2 years after THA. *Acta Orthop* 2010;81(2):183-8. doi: 10.3109/17453671003793204 [published Online First: 2010/04/07]
5. Vissers MM, Bussmann JB, Verhaar JA, et al. Recovery of physical functioning after total hip arthroplasty: systematic review and meta-analysis of the literature. *Phys Ther* 2011;91(5):615-29. doi: 10.2522/ptj.20100201 [published Online First: 2011/03/12]
6. Beswick AD, Wylde V, Gooberman-Hill R, et al. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;2(1):e000435. doi: 10.1136/bmjopen-2011-000435 [published Online First: 2012/02/24]
7. Fernandes L, Hagen KB, Bijlsma JW, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Ann Rheum Dis* 2013;72(7):1125-35. doi: 10.1136/annrheumdis-2012-202745 [published Online First: 2013/04/19]
8. Danish Health Authority. National Clinical Guideline on Hip Osteoarthritis - Non-surgical Treatment and Rehabilitation Following Total Hip Arthroplasty. 2nd ed: Danish Health Authority, 2021.
9. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;16(2):137-62. doi: 10.1016/j.joca.2007.12.013 [published Online First: 2008/02/19]
10. Brosseau L, Wells GA, Pugh AG, et al. Ottawa Panel evidence-based clinical practice guidelines for therapeutic exercise in the management of hip osteoarthritis. *Clin Rehabil* 2016;30(10):935-46. doi: 10.1177/0269215515606198 [published Online First: 2015/09/25]
11. van Doormaal MCM, Meerhoff GA, Vliet Vlieland TPM, et al. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. *Musculoskeletal Care* 2020 doi: 10.1002/msc.1492 [published Online First: 2020/07/10]
12. Goh SL, Persson MSM, Stocks J, et al. Relative Efficacy of Different Exercises for Pain, Function, Performance and Quality of Life in Knee and Hip Osteoarthritis: Systematic Review and Network Meta-Analysis. *Sports Med* 2019;49(5):743-61. doi: 10.1007/s40279-019-01082-0 [published Online First: 2019/03/05]
13. Hansen S, Mikkelsen LR, Overgaard S, et al. Effectiveness of supervised resistance training for patients with hip osteoarthritis - a systematic review. *Dan Med J* 2020;67(6) [published Online First: 2020/08/04]
14. Skoffer B, Dalgas U, Mechlenburg I. Progressive resistance training before and after total hip and knee arthroplasty: a systematic review. *Clin Rehabil* 2015;29(1):14-29. doi: 10.1177/0269215514537093 [published Online First: 2014/07/06]

15. Hermann A, Holsgaard-Larsen A, Zerahn B, et al. Preoperative progressive explosive-type resistance training is feasible and effective in patients with hip osteoarthritis scheduled for total hip arthroplasty--a randomized controlled trial. *Osteoarthritis Cartilage* 2016;24(1):91-8. doi: 10.1016/j.joca.2015.07.030 [published Online First: 2015/08/19]
16. Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *Jama* 2017;318(23):2337-43. doi: 10.1001/jama.2017.18556 [published Online First: 2017/12/21]
17. Chan AW, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *Bmj* 2013;346:e7586. doi: 10.1136/bmj.e7586 [published Online First: 2013/01/11]
18. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012;10(1):28-55. doi: 10.1016/j.ijsu.2011.10.001 [published Online First: 2011/11/01]
19. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42(2):377-81. doi: 10.1016/j.jbi.2008.08.010 [published Online First: 2008/10/22]
20. Beard DJ, Harris K, Dawson J, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68(1):73-9. doi: 10.1016/j.jclinepi.2014.08.009 [published Online First: 2014/12/03]
21. Costa ML, Achten J, Parsons NR, et al. Total hip arthroplasty versus resurfacing arthroplasty in the treatment of patients with arthritis of the hip joint: single centre, parallel group, assessor blinded, randomised controlled trial. *Bmj* 2012;344:e2147. doi: 10.1136/bmj.e2147 [published Online First: 2012/04/21]
22. Judge A, Arden NK, Price A, et al. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? *J Bone Joint Surg Br* 2011;93(12):1660-4. doi: 10.1302/0301-620x.93b12.27046 [published Online First: 2011/12/14]
23. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *The Journal of bone and joint surgery British volume* 2007;89(8):1010-4. doi: 10.1302/0301-620x.89b8.19424 [published Online First: 2007/09/06]
24. Skou ST, Roos EM, Laursen MB, et al. Total knee replacement plus physical and medical therapy or treatment with physical and medical therapy alone: a randomised controlled trial in patients with knee osteoarthritis (the MEDIC-study). *BMC Musculoskelet Disord* 2012;13:67. doi: 10.1186/1471-2474-13-67 [published Online First: 2012/05/11]
25. Detry MA, Lewis RJ. The intention-to-treat principle: how to assess the true effect of choosing a medical treatment. *Jama* 2014;312(1):85-6. doi: 10.1001/jama.2014.7523 [published Online First: 2014/07/25]
26. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012;367(14):1355-60. doi: 10.1056/NEJMsr1203730 [published Online First: 2012/10/05]
27. Dawson J, Fitzpatrick R, Carr A, et al. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78(2):185-90. [published Online First: 1996/03/01]
28. Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas* 2016;7:101-8. doi: 10.2147/prom.S97774 [published Online First: 2016/08/16]

29. Paulsen A, Odgaard A, Overgaard S. Translation, cross-cultural adaptation and validation of the Danish version of the Oxford hip score: Assessed against generic and disease-specific questionnaires. *Bone Joint Res* 2012;1(9):225-33. doi: 10.1302/2046-3758.19.2000076 [published Online First: 2013/04/24]
30. Pham T, van der Heijde D, Altman RD, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis Cartilage* 2004;12(5):389-99. doi: 10.1016/j.joca.2004.02.001 [published Online First: 2004/04/20]
31. Harris KK, Price AJ, Beard DJ, et al. Can pain and function be distinguished in the Oxford Hip Score in a meaningful way? : an exploratory and confirmatory factor analysis. *Bone Joint Res* 2014;3(11):305-9. doi: 10.1302/2046-3758.311.2000313 [published Online First: 2014/11/05]
32. de Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage* 2007;15(1):104-9. doi: 10.1016/j.joca.2006.06.014 [published Online First: 2006/08/08]
33. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol* 2003;32(1):46-51. [published Online First: 2003/03/15]
34. Nilsdotter AK, Lohmander LS, Klassbo M, et al. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10. doi: 10.1186/1471-2474-4-10 [published Online First: 2003/06/05]
35. Thorborg K, Roos EM, Bartels EM, et al. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med* 2010;44(16):1186-96. doi: 10.1136/bjsm.2009.060889 [published Online First: 2009/08/12]
36. Amstutz HC, Thomas BJ, Jinnah R, et al. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. *The Journal of bone and joint surgery American volume* 1984;66(2):228-41. [published Online First: 1984/02/01]
37. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clin Orthop Relat Res* 2009;467(4):958-65. doi: 10.1007/s11999-008-0358-5 [published Online First: 2008/07/01]
38. Terwee CB, Bouwmeester W, van Elsland SL, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage* 2011;19(6):620-33. doi: 10.1016/j.joca.2011.01.002 [published Online First: 2011/01/22]
39. Wright AA, Cook CE, Baxter GD, et al. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. *J Orthop Sports Phys Ther* 2011;41(5):319-27. doi: 10.2519/jospt.2011.3515 [published Online First: 2011/02/22]
40. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21(8):1042-52. doi: 10.1016/j.joca.2013.05.002 [published Online First: 2013/05/18]
41. Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. *Physiother Res Int* 2008;13(3):141-52. doi: 10.1002/pri.411 [published Online First: 2008/08/13]

42. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Res Q Exerc Sport* 1999;70(2):113-9. doi: 10.1080/02701367.1999.10608028 [published Online First: 1999/06/25]
43. Christensen R, Bliddal H, Henriksen M. Enhancing the reporting and transparency of rheumatology research: a guide to reporting guidelines. *Arthritis Res Ther* 2013;15(1):109. doi: 10.1186/ar4145 [published Online First: 2013/03/02]
44. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. *J Clin Epidemiol* 2021;134:174-77. doi: 10.1016/j.jclinepi.2021.03.009 [published Online First: 2021/05/22]
45. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *Journal of clinical epidemiology* 2014;67(7):769-72. doi: 10.1016/j.jclinepi.2013.11.011 [published Online First: 2014/02/25]
46. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. *Jama* 2016;315(4):407-8. doi: 10.1001/jama.2015.19394 [published Online First: 2016/01/28]
47. White IR, Horton NJ, Carpenter J, et al. Strategy for intention to treat analysis in randomised trials with missing outcome data. *Bmj* 2011;342:d40. doi: 10.1136/bmj.d40 [published Online First: 2011/02/09]
48. International Conference on Harmonisation Expert Working Group. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use: ICH harmonised tripartite guideline. Guideline for Good Clinical Practice E6 (R1). 1996 [Available from: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf accessed June 20, 2019.
49. Danish Health Authority. Sundhedsstyrelsens udmeldinger om alkohol 2021 [Available from: https://www.sst.dk/da/viden/alkohol/alkoholforebyggelse/sundhedsstyrelsens-udmeldinger-om-alkohol accessed 14-05-2021.

Table 1. Baseline Characteristics of the Patients in the ITT population.*

	Total Hip Arthroplasty (N=??)	Progressive Resistance Training (N=??)
Female sex — no. (%)		
Age — yr		
Height — m		
Weight — kg		
Body-mass index — kg/m^2		
Education level beyond high school† — no. (%)		
Employment		
Employed for wages		
Self-employed		
Sick leave		
Retired		
Other		
Substance use — no. (%)		
Current smoker		
Alcohol consumption above recommendations†		
Index hip right — no. (%)		
Duration of hip symptoms — yr		
Previous total hip arthroplasty — no. (%)		
Previous total knee arthroplasty — no. (%)		
Used pain medication in the past week due to hip-related pain — no. (%)		
Paracetamol		
Ibuprofen		
Morphine or opioids		
Other		
OHS score‡ — 0 to 48		
HOOS subscale scores§ — 0 to 100		
Pain		
Symptoms		
Function in activities of daily living		
Hip-related quality of life		
Function in sports and recreation		
UCLA activity score¶ — 1 to 10		
Walking speed in the 40 meter fast-paced walk test — m/s		
Repetitions in the 30-second chair stand test — no.		

* Plus–minus values are mean ±SD unless otherwise indicated.

† The Danish Health Authority guideline recommends alcohol consumption lower than 7 units per week for females and 14 units per week for males to have a low risk of developing diseases.[49]

‡ The Oxford Hip Score (OHS) ranges from 0 to 48, with higher scores indicating better disease status.

§ For all five subscales, the Hip disability and Osteoarthritis Outcome Score (HOOS) ranges from 0 to 100, with higher scores indicating better disease status.

¶ The University of California Los Angeles (UCLA) Activity Score ranges from 1 to 10, with higher scores indicating greater physical activity level.

Table 2. Change from Baseline in Primary and Key Secondary Outcomes at 6 months in the Intention-to-Treat Population.*

Outcome	Change from Baseline to 6 Months		Between-Group Difference in Mean Improvement	
	Total Hip Arthroplasty	Progressive Resistance Training	Difference in LSMeans (95% CI)	P-Value
Primary Outcome				
OHS score — 0 to 48				
Key Secondary Outcomes				
HOOS subscale scores — 0 to 100				
Pain				
Symptoms				
Function in activities of daily living				
Hip-related quality of life				
Function in sports and recreation				
UCLA activity score — 1 to 10				
Walking speed in the 40 meter fast-paced walk test — m/s				
Repetitions in the 30-second chair stand test — no.				
Response to Treatment				
OHS minimal important change criteria† — no. (%)				
OMERACT-OARSI criteria‡ — no. (%)				

* All analyses will be based on the Intention-To-Treat population: Using repeated measures linear mixed effects models (with no imputation for missing data); Estimates are least squares means (LSMeans) and standard errors with difference between groups reported with 95% confidence intervals.

† Patients will be classified as having a clinically relevant change if the OHS score improves by 8 points or more.[20]

‡According to the criteria of the Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative (OMERACT–OARSI),[30] patients will be classified as having a response if the OHS pain or function subscale score improves by 50% or more and by 20 or more points or if two of the following three findings is observed: an improvement in the OHS pain score by 20% or more and by 10 or more points on the subscale, an improvement in the OHS function score by 20% or more and by 10 or more points on the subscale, or an improvement in the score on the patient's OHS score by 20% or more and by 10 or more points on the scale. The OHS pain and function subscales ranges from 0 to 100, with higher scores indicating better disease status.[31]

Table 3. Serious Adverse Events at 6 Months in All Patients Assigned to Treatment.*

Events	Total Hip Arthroplasty	Progressive Resistance Training
Serious adverse event — no. (%)		
Musculoskeletal		
Deep infection		
Hip dislocation		
Femoral fracture		
Aseptic loosening		
Cardiovascular		
Vascular injury		
Pulmonary embolism		
Deep venous thrombosis		
Acute myocardial infarction		
Stroke		
Nervous system		
Nerve injury		
Deaths		
Discontinuation due to serious adverse event(s) — no. (%)		

* This table includes all serious adverse events that occurred during the 6-month study period, but which did not necessarily have a causal relationship with the treatment administered. An adverse event was classified as serious if it was fatal or life-threatening, required or prolonged inpatient hospitalization, was disabling, resulted in (a congenital anomaly or birth defect), or required medical or surgical intervention to prevent permanent impairment or damage.

FIGURE 1

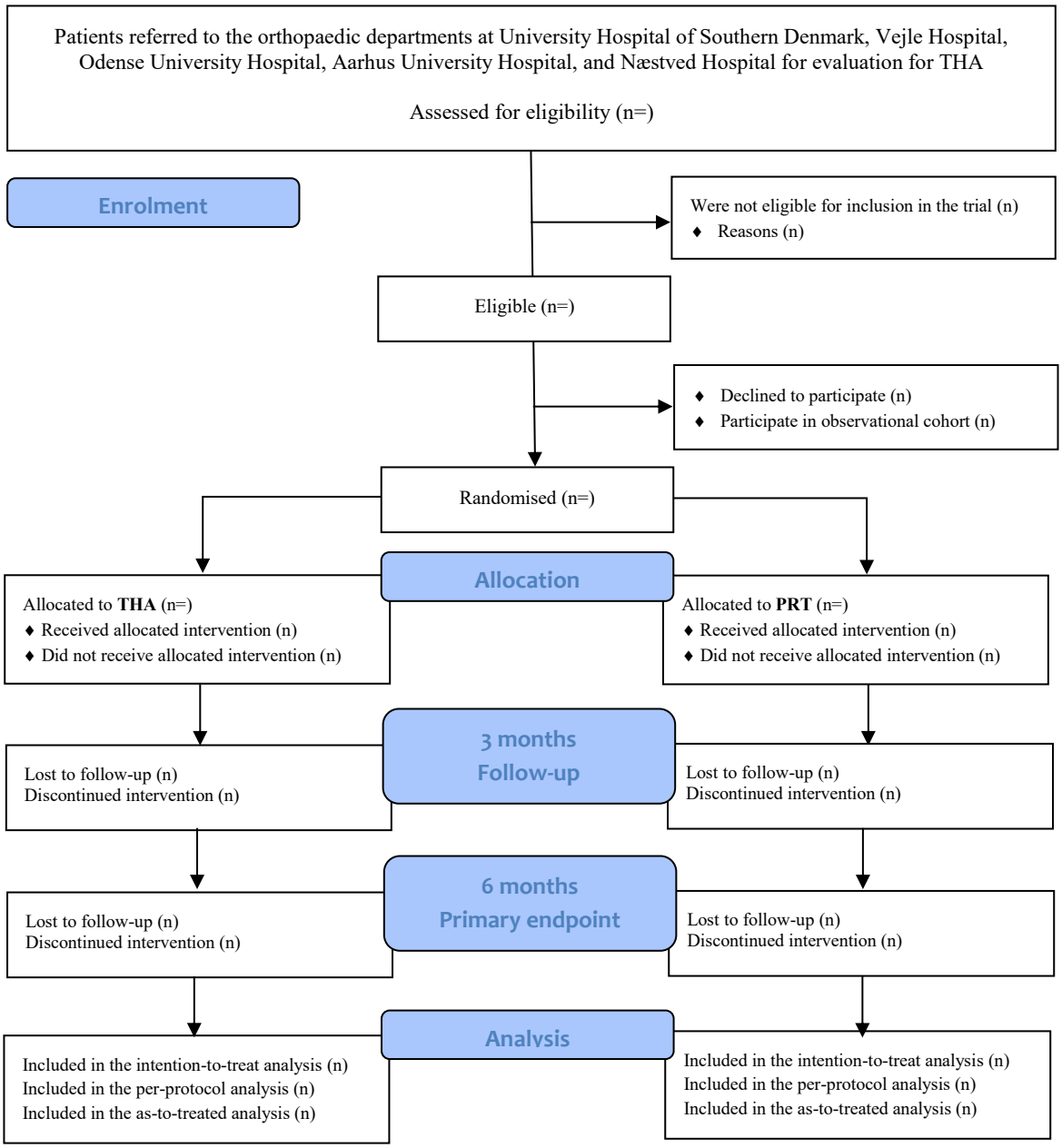


Figure 1. CONSORT flow-chart. Total hip arthroplasty (THA), Progressive resistance training (PRT).

FIGURE 2

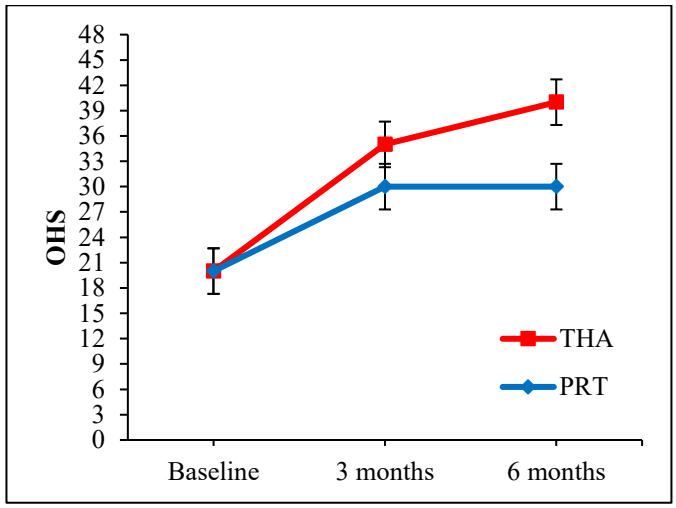


Figure 2. Trajectory in the Oxford Hip Score (OHS) in the total hip arthroplasty (THA) group and progressive resistance training (PRT) group at baseline, 3, and 6 (primary end-point) months after initiating the treatment. Values are mean (95% confidence intervals). This figure is depicted with anticipated changes from baseline to 6-months follow-up.

Appendix Table 1. Change from Baseline in Primary and Key Secondary Outcomes at 6 months in the Per-Protocol Population.*

Outcome	Change from Baseline to 6 Months		Between-Group Difference in Mean Improvement	
	Total Hip Arthroplasty†	Progressive Resistance Training‡	Difference in LSMeans (95% CI)	P-Value
Primary Outcome				
OHS score — 0 to 48				
Key Secondary Outcomes				
HOOS subscale scores — 0 to 100				
Pain				
Symptoms				
Function in activities of daily living				
Hip-related quality of life				
Function in sports and recreation				
UCLA activity score — 1 to 10				
Walking speed in the 40 meter fast-paced walk test — m/s				
Repetitions in the 30-second chair stand test — no.				

* All analyses will be based on the per protocol population: Using repeated measures linear mixed effects models (with no imputation for missing data); Estimates are least squares means (LSMeans) and standard errors with difference between groups reported with 95% confidence intervals.

† The Per-Protocol Population included N=?? in the Total Hip Arhtroplasty group.

‡ The Per-Protocol Population included N=?? in the Progressive Resistance Group group.

Appendix Table 2. Change from Baseline in Primary and Key Secondary Outcomes at 6 months in the Intention-to-Treat Population using a single-step non-responder imputation.*

Outcome	Change from Baseline to 6 Months		Between-Group Difference in Mean Improvement	
	Total Hip Arthroplasty	Progressive Resistance Training	Difference in LSMeans (95% CI)	P-Value
Primary Outcome				
OHS score — 0 to 48				
Key Secondary Outcomes				
HOOS subscale scores — 0 to 100				
Pain				
Symptoms				
Function in activities of daily living				
Hip-related quality of life				
Function in sports and recreation				
UCLA activity score — 1 to 10				
Walking speed in the 40 meter fast-paced walk test — m/s				
Repetitions in the 30-second chair stand test — no.				

* All analyses will be based on the Intention-to-Treat population: Using repeated measures linear mixed effects models (missing data is replaced by baseline observation carried forward); Estimates are least squares means (LSMean) and standard errors with difference between groups reported with 95% confidence intervals.

Appendix Table 3. Change from Baseline in Primary and Key Secondary Outcomes at 6 months in the As-Treated Population.*

Outcome	**Change from Baseline to 6 Months**				
	Total Hip Arthroplasty†	Progressive Resistance Training‡	Total Hip Arthroplasty and no surgery§	Progressive Resistance Training and surgery¶	P-Value (main effect)
Primary Outcome					
OHS score — 0 to 48					
Key Secondary Outcomes					
HOOS subscale scores — 0 to 100					
Pain					
Symptoms					
Function in activities of daily living					
Hip-related quality of life					
Function in sports and recreation					
UCLA activity score — 1 to 10					
Walking speed in the 40 meter fast-paced walk test — m/s					
Repetitions in the 30-second chair stand test — no.					

* All analyses will be based on the As Treated Population: Using repeated measures linear mixed effects models; Estimates are least squares means and standard errors.

An as-treated analysis will be performed based on their patients' adherence to the randomised treatment expecting four groups:

† Patients randomised to Total Hip Arthroplasty undergoing surgery included N=??.

‡ Patients randomised to Progressive Resistance Training without undergoing surgery in the follow-up period included N=??.

§ Patients randomised to Total Hip Arthroplasty but declined surgery post randomisation included N=??.

¶ Patients randomised to Progressive Resistance Training undergoing surgery during the follow-up period N=??.