Statistical Analysis Plan (SAP)

Trial: The effects of adding aerobic physical activity to strengthening exercise on hip osteoarthritis symptoms - the PHOENIX trial

Contents

Section 1. Administrative Information

1. Title

The effects of adding aerobic physical activity to strengthening exercise on hip osteoarthritis symptoms – the PHOENIX trial.

2. Trial registration

Prospectively registered (Australian New Zealand Clinical Trials Registry Trial Id: ACTRN12619001297112, 20/09/2019).

3. SAP version

Version: 1.0 Date: 16/11/2023

4. Protocol Version

This document has been written based on information contained in the PHOENIX study protocol version 1.7 dated 8/12/2021. The protocol was published as follows:

Hall M, Allison K, Hinman RS, Bennell KL, Spiers L, Knox G, Plinsinga M, Klyne DM, McManus F, Lamb KE, da Costa R, Murphy NJ, Dobson F. Effects of strengthening and physical activity compared to strengthening only in people with hip osteoarthritis: protocol for the PHOENIX randomised controlled trial. *BMC Musculoskeletal Disorders* 2022; 23:361.

5. SAP Revisions

Not applicable

6. Names and affiliations

Document prepared by A/Professor Michelle Hall, A/Professor Fiona Dobson, Dr Kim Allison, Ms Gabrielle Knox, A/Professor Karen Lamb, Dr Anurika De Silva and Ms Fiona McManus.

Emails: M Hall halm@unimelb.edu.au; F Dobson fdobson@unimelb.edu.au; G Knox gabby.knox@unimelb.edu.au; K Lamb klamb@unimelb.edu.au, A De Silva anurika.de@unimelb.edu.au; F McManus fmcmanus@unimelb.edu.au

Signatures:

Signature of statistician responsible (Ms Fiona McManus): *Date: 14/11/2023*

Signature of chief investigator (A/Prof Michelle Hall): *Date: 16/11/2023*

Section 2: Introduction

7. Background and rationale

Osteoarthritis (OA) is the 11th leading cause of disability worldwide [1] and hip OA affects one in four adults over their lifetime [2]. In Australia, arthritis-related health care costs exceed $2.1 billion AUD annually, of which OA is the largest contributor [3]. The greatest driver of health care costs for hip OA is joint replacement surgery. The lifetime risk of hip replacement for hip OA in the population is up to 12.6% [4]. Treatments that reduce symptoms and delay the need for joint replacement are critical. Current OA clinical guidelines emphasise that non-drug, non-surgical strategies [5-7] are the core management strategies for hip OA and should be offered prior to consideration for surgical management.

A 2017 meta-analysis of land-based exercise randomised controlled trials (RCTs) in hip OA identified 12 eligible RCTs and reported small-to-modest beneficial effects of exercise on pain (effect size -0.24, 95%CI: -0.42, -0.06) and physical function (effect size -0.34, 95%CI: -0.50, -0.18) compared to no exercise [8, 9]. Of note, all trials in the systematic review evaluated lower-limb muscle strengthening exercise, while only 3 investigated aerobic physical activity. Thus, current advice advocating exercise for hip OA is predominantly based on lower-limb strengthening interventions, which may account for the reported small-to-modest beneficial effects of land-based exercise on hip OA symptoms. Hip and thigh muscle weakness is widely established in people with hip OA [10]. However, muscle strengthening exercise in isolation likely inadequately alleviates pain and physical dysfunction in many people with hip OA. In 195 people with hip OA, we found very large and unfeasible increases in muscle strength are probably required to achieve clinically meaningful improvements in symptoms for many patients [11] – highlighting the limitation of muscle strengthening alone to improve hip OA symptoms.

Many people with hip OA do not meet aerobic physical activity guidelines [12], defined by The World Health Organisation as at least 150 minutes of moderate-intensity aerobic physical activity or at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week [13]. It is therefore perhaps unsurprising that people with hip OA have lower cardiovascular fitness [14], higher levels of depression and stress [15], and more often have overweight or obesity [16] compared to those without hip OA. Of concern, cardiovascular disease [17], depressive symptoms [18, 19], and obesity [18, 19] are associated with functional decline in hip and knee OA populations. Aerobic exercise, with or without strengthening exercise, improves cardiovascular fitness and psychological well-being and reduces fat mass compared to strengthening exercise alone. In a hip OA RCT, aerobic exercise had greater beneficial effects on cardiovascular fitness, overall mental health and self-efficacy related to OA symptoms compared to strengthening alone [20]. Moreover, evidence from clinical trials in healthy adults [21] and dieting adults with obesity [22], has demonstrated superior beneficial effects of a combined aerobic and strengthening exercise program on cardiovascular fitness [22], mental health [22] and fat mass [21], compared with strengthening exercise alone. However, there are no trials in hip OA evaluating the addition of aerobic exercise to a strengthening exercise program, compared to a strengthening program alone.

8. Objectives

Study objectives:

Primary objective: To determine if adding aerobic physical activity to a lower-limb strengthening exercise program leads to a greater reduction in hip pain and improved physical function when compared to a lower-limb strengthening exercise program alone at 3 months.

Secondary objective: To determine if adding aerobic physical activity to a lower-limb strengthening exercise program will have greater benefits for other outcomes (other pain measures, stiffness, global rating, quality of life, muscle strength, physical function, cardiorespiratory fitness) when compared to a lower-limb strengthening exercise program alone at 3 months.

Research hypotheses:

Primary alternative hypotheses:
Adding aerobic physical activity to a lower-limb strengthening exercise program will lead to greater reduction in hip pain and improved physical function when compared to a lower-limb strengthening exercise program at 3 months.

Secondary alternative hypothesis: Adding aerobic physical activity to a lower-limb strengthening exercise program will have greater benefits for other outcomes (other pain measures, stiffness, global rating, quality of life, muscle strength, physical function, cardiorespiratory fitness) when compared to a lower-limb strengthening exercise program alone at 3 months.

Interpretation of primary outcomes:

We will consider the addition of aerobic physical activity to be effective for *symptoms* if a difference in change between groups for the primary outcomes of pain and physical function over 3 months is found favouring the addition of aerobic physical activity to a lower-limb strengthening exercise program.

Section 3: Trial Methods

9. Trial design

The PHOENIX trial is designed as a two-group, superiority, randomised, placebo-controlled trial. Treatment allocation is a 1:1 ratio.

10. Randomisation

Prior to March, 2020, participants self-selected one of 8 study physiotherapists for face-to-face consultations. Due to restrictions related to the COVID-19 pandemic, from March, 2020 until study completion, participants were randomised to one of 8 study physiotherapists using a block randomisation list prepared by an independent statistician. If a physiotherapist was unavailable, participants were re-randomised to another available physiotherapist. Once allocated to a physiotherapist, participants were allocated to a treatment group using a randomisation schedule prepared by an independent biostatistician according to a 1:1 allocation in blocks of varying sizes stratified by physiotherapist. If a physiotherapist was unavailable following randomisation to treatment group, the participant was re-randomised to a physiotherapist and the physiotherapist recorded was the physiotherapist that conducted more than 50% of physiotherapy sessions. The randomisation schedule is stored on a password-protected website (REDCap™) maintained by a researcher not involved in either participant recruitment or administration of the outcome measures.

11. Sample size

Primary outcomes will be the 3-month change in (i) overall average hip pain over the last week measured using an 11-point numeric rating scale with terminal descriptors ‘no pain’ (score 0) and ‘worst pain possible’ (score 10); and (ii) physical function over the last week measured using the physical function subscale of the Western Ontario and McMaster Universities. The minimum clinically important difference to be detected in OA trials is a change in pain of 1.8 (out of 10 units)[23].

Based on data from our previous trial on exercise in hip OA [24, 25], we assumed between-participant standard deviations of 2.2 for pain and 13 units for physical function, and a baseline to 3-month correlation of 0.46 for pain and 0.40 for physical function respectively. Using analysis of covariance adjusted for baseline score, and to achieve 90% power with a 5% significance level, we required 25 participants per arm to detect the MCID in between-group change in pain and 83 per arm for physical function. Allowing for 15% attrition, we recruited 98 participants per arm for a total sample of 196 participants.

12. Framework

This trial uses a superiority hypothesis testing framework between groups for all outcomes.

13. Statistical Interim analyses and stopping guidance

Nil

14. Timing of final analysis

Final analysis will be performed after all participants have reached the 9-month timepoint and completed assessments.

15. Timing of outcome assessments

Table 4.6 in the study protocol details the timing of outcome assessments, the majority of which occur at baseline, 3- and 9-months.

Section 4: Statistical Principles

16. Level of statistical significance

All applicable statistical tests will be 2-sided and will be performed using a 5% significance level.

17. Description of any planned adjustment for multiplicity, and if so, including how the type 1 error is to be controlled

We have two primary outcomes, which are both equally important: one for hip pain and one for hip related physical function. We have multiple secondary outcomes. We will not adjust for co-primary outcomes nor multiple secondary outcomes but instead report all effect sizes, confidence intervals, and p values in order to let readers use their own judgment about the relative weight of the conclusions on the effect of adding aerobic physical activity for pain and physical function. This approach aligns with the usage of p-values favoured by the American Statistical Association[26]. As the sample size was based on a 5% significance level unadjusted for duel primary outcomes, the trial will only be declared successful if both primary outcomes are successful at 3 months.

18. Confidence intervals to be reported

All confidence intervals will be 95% confidence intervals.

19. Adherence and Protocol Deviations

The primary analysis will be based on the principle of intention-to-treat, whereby participants are included in the groups to which they were originally assigned, regardless of their adherence to their assigned treatments. Any protocol deviations (if they occur), including errors applying inclusion/exclusion criteria and/or administration of the wrong intervention will be summarised in trial results (patient flow diagram/text) by treatment group. Randomisation errors resulting from these errors will be handled according to recommendations [15].

In this trial, participants record the exercise type, dosage and intensity performed in their logbooks. Participant exercise logbooks are used to determine if 3 strengthening exercises (from the first consultation) and 4–6 strengthening exercises (from the second consultation onwards) at least at a “very hard” intensity (Rating of

Perceived Exertion, RPE) were performed, at least three times per week. Adherence for the strengthening group is considered satisfactory if participants self-report completing strengthening exercises "very hard" at least 10 weeks of the 3 months. Adherence for the combined aerobic physical activity and strengthening exercise group, is considered satisfactory if participants self-report completing strengthening exercises "very hard" and self-report completing the number of minutes working at least "somewhat hard" (RPE) as set by the physiotherapist in at least 10 weeks of the 3-month program.

Participants in the combined aerobic and strengthening exercise group will be classified as 'fully adherent' if they self-report completing the number of minutes working at least "somewhat hard" (RPE) as set by the physiotherapist in at least 8 weeks of the 3-month program, and all other participants in the combined aerobic and strengthening exercise group will be classified as 'non-adherent'. If a participant does not provide a self-report, including if they provide less than 8 weeks of self-reports, non-adherence will be assumed by performing no imputation for the missing adherence variable.

20. Analysis Populations

The primary analysis will be based on the principle of intention-to-treat, as described in Section 19. A complier average causal effects analysis will also be undertaken to determine the treatment effect at 3 months among those that were fully adherent to the aerobic component of the intervention (as defined in Section 19).

Section 5: Trial Population

21. Screening Data

Screening data will be collected and summarized. A CONSORT flow diagram will be used [16]. The following summaries will be presented in text and/or flow diagram: time frame for recruitment, the number of patients screened, the number of patients recruited, the number of screened patients not recruited, and the reasons for non-recruitment.

22. Eligibility

Trial inclusion and exclusion criteria are described in section 5.2 of the trial protocol. Reasons for exclusion will be summarized in the CONSORT [16] flow diagram.

23. Recruitment

A CONSORT flow diagram [16] will be used to describe the number of people enrolled, randomized, allocated to each treatment group, lost to follow up (including reasons) and analysed.

24. Withdrawal/follow-up

If a participant withdraws from the study, the nature, timing of and reasons for withdrawal will be described (provided the participant responds to contact made by the research team). Any data provided up to the point of withdrawal will be analysed in accordance with intention to treat analyses, unless the participant specifically requests to withdraw their data from the study. Losses to follow-up (including reasons) will be summarised in the CONSORT flow diagram by treatment group.

25. Baseline characteristics

Baseline characteristics will be summarised by treatment group and presented in a table:
- Age
- Gender
- Height, body mass, body mass index

- Education level
- Current employment status
- Duration of hip OA symptoms
- Unilateral or bilateral hip symptoms (laterality of symptoms)
- Co-morbidities
- Radiographic disease severity using the Kellgren & Lawrence (KL) scale
- Expectation of treatment outcome
- Co-intervention use
- Pain type
- Sleep quality

Baseline characteristics will be summarised as appropriate (means and standard deviations for continuous variables that are approximately normally distributed, medians and interquartile ranges for other continuous variables, counts and percentages for categorical variables). Tests of statistical significance will not be undertaken for comparing baseline characteristics of treatment groups; rather the clinical importance of any imbalance will be noted.

An appendix table will provide summaries of baseline characteristics and baseline levels of primary and secondary outcomes between two groups: those participants who provide both primary outcomes at 3 months, and those participants who are missing one or both primary outcomes. Tests of statistical significance will not be undertaken to compare the two groups; rather, the clinical importance of any imbalance will be noted.

Section 6: Analysis

26. Outcome definitions

Co-primary outcomes:

- Change in severity of hip pain: Average overall hip pain in the past week is self-assessed using a 11-point numeric rating scale with terminal descriptors of ‘no pain’ (score 0) and ‘worst pain possible’ (score 10). Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

- Change in difficulty with physical function: Difficulty with physical function in the past 48 hours is self-assessed using 17-items on the physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (Likert version 3.1) with terminal descriptors of ‘no difficulty’ (score 0) and ‘extreme difficulty’ (score 4). Scores range between 0 to 68, with higher scores indicating greater difficulty with physical dysfunction. Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

Secondary outcomes:

- Change in severity of hip pain while walking: Average overall knee pain while walking in the past week is self-assessed using a 11-point numeric rating scale with terminal descriptors of ‘no pain’ (score 0) and ‘worst pain possible’ (score 10). Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

- Change in each of pain and stiffness in the last week: Using subscales of the Western Ontario and McMaster Universities Osteoarthritis Index (Likert version 3.1), pain (5-items); ii) stiffness (2-items) will be self-assessed. Responses are provided on a 5-point scales. Scores will be calculated for pain and stiffness and range from 0 to 20, and 0 to 8 respectively with higher scores indicating greater symptoms severity. Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

- Global improvement at 3 and 9 months: Global improvement in pain and physical function will be scored using a 7-point global rating of change Likert scale with response options ranging from “much worse” to

"much better" when compared to baseline (3 months) and over the past 6 months (9 months). Participants indicating they are "moderately better" or "much better" will be classified as improved. All other respondents will be classified as not improved.

- Change in health-related quality of life: The AQoL questionnaire (version AQoL-6D) measures health-related quality of life. This is a 20-item questionnaire and scores range from -0.04 to 1.00 with 1.00 indicating full health-related quality of life. Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

- Change in physical function using patient specific function scale: At baseline, participants indicate up to five activities of daily living in which performance was limited in the previous week and rate these on an 11-point NRS ranging from 0-10, where 0 indicates 'unable to perform' and 10 indicates 'able to perform at the same level as before injury or problem'. Total score is the sum of the ratings for each activity divided by the number of activities, with higher scores indicating better function [39]. At 3 and 9 months, participants are asked about the activities they listed at baseline and provide a rating on the same scale. Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.

- Change in muscle strength: Maximal normalised isometric hip abduction, hip extension and knee extensor strength (peak torque, Nm/kg) will be assessed at baseline and 3 months [41, 42]. The sum of hip abduction, hip extensor and knee extensor strength is calculated and will be reported as an overall strength score. Change score at 3 months (primary time point) will be calculated as baseline minus follow-up.

- Change in physical performance measured using three separate tests: At baseline and 3 months, physical performance is measured using the Osteoarthritis Research Society International set of performance-based measures of physical function, including the 30-second sit-to-stand test, 40-meter fast pace walk test and 6-step stair climb test [43, 44]. For the 30-second sit-to-stand test, following a practice trial, the number of times participants can rise to a full standing position from sitting and return to sitting, with arms crossed and held against the chest, within 30 seconds is counted. A greater number indicates better performance. For the 40m fast pace walk test, participants are timed (seconds) to determine how long it takes them to walk 40m in their usual footwear with the instruction 'walk as quickly as you can without overexerting yourself'. This is performed once and the 40m time recorded, with less time indicating better performance. For the 6-step stair climb test, the time (seconds) to walk up and down six 17.5 cm high steps as quickly as possible using a handrail if preferred is recorded, with less time indicating better performance. Change scores for each of the 30-second sit-to-stand test, 40-meter fast pace walk test and 6-step stair climb test at 3 months (primary time point) will be calculated as baseline minus follow-up.

- Change in cardiorespiratory fitness: At baseline and 3 months, a submaximal exercise test is performed on a cycle ergometer (Corival, Lode B.V. Groningen, The Netherlands), and submaximal cardiorespiratory fitness is measured by breath-by-breath indirect calorimetry using a Vmax Encore metabolic cart (Carefusion, San Diego, CA, USA). Procedures are adjusted from standard fitness testing protocols [45] to suit participant clinical presentation. Cardiorespirator fitness will be expressed as the VO_2 in milliliters per kilogram at which the respiratory exchange ratio is 1.000 [27]. Change scores at 3 months (primary time point) will be calculated as baseline minus follow-up.

27. Analysis methods

This statistical analysis plan will be published on the Centre for Health, Exercise and Sports Medicine's website while all authors for this analysis plan are blinded. A biostatistician (FM, supervised by ADS and KEL) will analyse data blinded to group name. Main comparative analyses between groups will be performed using intention-to-treat. If more than 5% of primary outcomes are missing at 3 months (primary time point), multiple imputation will be applied as the primary analysis. Differences in mean change (baseline minus follow-up) at each timepoint (3-months and 9-months) in each primary outcome (pain and physical function) will be compared between groups separately using linear mixed-effects modelling adjusted for the relevant outcome at baseline and

the stratifying variable, physiotherapist, with random effects to account for clustering by participants. Models will include factors representing intervention, time and the intervention by time interaction. For the primary hypotheses, the absolute difference in mean change from baseline between groups will be estimated (including two-sided 95% confidence intervals) at 3 months (primary time point). We will consider a 5% significance level for each of the primary outcomes at the primary timepoint, although emphasis will be placed on the clinical meaning of effect sizes and confidence intervals when interpreting the results. Standard diagnostic plots will be used to check model assumptions.

Similar analyses to that conducted for the primary outcomes will be conducted for continuous secondary outcomes measured at baseline, 3 and 9 months. For continuous secondary outcomes measured only at baseline and 3 months, differences in mean change (baseline minus follow-up) in each outcome will be compared between groups separately using multiple linear regression adjusted for the relevant outcome at baseline and physiotherapist. Improvement based on global change at each timepoint (3 and 9 months) will be compared between groups using risk ratios and risk differences, by fitting mixed-effects logistic regression models, adjusted for physiotherapist, with random effects to account for clustering by participants. All confidence intervals will be 95% confidence intervals. Standard diagnostic plots will be used to check model assumptions.

Adherence, adverse events and other measures will be summarised by intervention group (and 3- and 9-month timepoint, if relevant), and presented in tables, unless otherwise stated:

- Number of consultations with physiotherapist
- Number of strength sessions at least “very hard” (Rating of Perceived Exertion, RPE)
- Number of minutes working at least “somewhat hard” (RPE) for those in the aerobic group
- Satisfactory adherence to the strengthening component as defined in Section 19
- Satisfactory adherence to the combined aerobic and strengthening component as defined in Section 19 for those in the aerobic group
- Full adherence to the aerobic component as defined in Section 19 for those in the aerobic group
- Adverse events
- Co-intervention use
- Treatment fidelity
- Overall satisfaction with the exercise program
- Therapeutic alliance using the Working Alliance Inventory Short Form (at 6 weeks).^

^Summarized in text or table.

Adherence, adverse events and other measures will be summarised as means and standard deviations for continuous variables and counts and percentages for categorical variables.

28. Statistical Methods – adjustment for covariates

As described above, analyses for all outcomes will be conducted adjusting for the relevant outcome at baseline, if available, and the stratifying variable, physiotherapist, as covariates. Where the outcome was measured at 3 and 9 months with or without a measure at baseline, a term for time and an interaction between time and intervention will also be included as covariates, and random effects will be included for participants.

29. Statistical Methods – sensitivity analyses

A sensitivity analysis will estimate treatment effects on the primary outcomes at 3 months assuming full adherence to the aerobic component of the intervention (full adherence defined as in Section 19). This analysis will assume that adherence to the strengthening exercise component is the same in both arms. Complier average causal effects will be estimated using an instrumental variables approach (where randomisation is the instrument for adherence). Two-stage least squares models will be fit, and complier average causal effects reported with 95% confidence intervals and p-values. This analysis will be performed unblinded to group name after all other analyses are complete.

If multiple imputation is required to handle missing data, complete-case analyses of primary and secondary outcomes will also be conducted in sensitivity analyses.

30. Statistical Methods – subgroup analyses

We will also conduct planned exploratory analyses to investigate potential moderators that could influence response to treatment for the two primary outcomes at 3 months. Pre-identified potential moderators are body mass index (BMI), pain type, sleep quality and radiographic disease severity. To assess the moderation of the effect of randomised treatment group by potential moderators, an interaction term between randomised group and the potential moderator, as well as terms for the randomised group and the potential moderator, will be included in multiple linear regression models, adjusted for the stratifying variable and the relevant outcome at baseline. Results will be calculated as the estimated mean effect (95% confidence interval) on each of the primary outcomes at 3 months of a one-unit increase in each potential continuous moderator, body mass index, pain type or sleep quality, for each group. Radiographic disease severity will be dichotomised (Kellgren-Lawrence grade <2 versus ≥ 2) and results calculated as the estimated mean between-group difference (95% confidence interval) on each of the primary outcomes at 3 months for each disease severity category.

The hypothesis for each moderator analysis is below:

- BMI: The symptom-relieving benefits of adding aerobic physical activity to lower-limb muscle strengthening compared to lower-limb strengthening only will be greater in those with a higher BMI compared to lower BMI.

- Pain type: The symptom-relieving benefits of adding aerobic physical activity to lower-limb muscle strengthening compared to lower-limb strengthening only will be greater in those with more probable neuropathic symptoms (higher PainDETECT scores) compared to those with more nociceptive symptoms (lower PainDETECT scores).

- Sleep quality: The symptom-relieving benefits of adding aerobic physical activity to lower-limb muscle strengthening compared to lower-limb strengthening only will be greater in those with poor sleep (lower Pittsburgh Sleep Quality Index scores) compared to those with good sleep (higher Pittsburgh Sleep Quality Index scores).

- Radiographic disease severity: The symptom-relieving benefits of adding aerobic physical activity to lower-limb muscle strengthening compared to lower-limb strengthening only will be similar between those with a Kellgren-Lawrence grade of ≥ 2 and those with a grade of <2.

31. Missing data reporting and assumptions/statistical methods to handle missing data

If more than 5% of participants have at least one primary outcome missing at 3 months, multiple imputation will be applied. The number of imputed datasets will be approximately equal to the proportion of participants missing the most primary outcome data, whichever primary outcome this is, with a minimum of 10 imputations conducted. Missing baseline characteristics will be imputed using single mean imputation. Missing outcome values will be imputed separately by treatment group, using chained equations and predictive mean matching, using the five nearest neighbours. Imputation models will include baseline levels of outcomes, age, sex, body mass index, current employment status, education level, laterality of hip symptoms, duration of hip symptoms in the study hip, co-morbid conditions, radiographic disease severity of hip osteoarthritis, expectation of treatment outcome, pain type and sleep quality. Initially, imputation models will include all primary and secondary outcomes together, with outcomes broken into subsets if imputation models do not converge. Imputed datasets will be compared to complete data using density plots for continuous outcomes and descriptive tabulations for the binary outcome.
To assess the potential impact of the violation of the missing-at-random assumption on conclusions for the

primary outcomes, a pattern-mixture approach (as in White et al [28]) will be applied. We will explore the impact of the violation of the missing-at-random assumption if the assumption was violated in both groups, or in one group only.

32. Additional Analyses

N/A

33. Harms

The number (and percentage) of patients experiencing adverse events will be presented for each treatment group and the nature of the event(s) described. Adverse events are defined as any untoward medical occurrence in a trial participant which does not necessarily have a causal relationship with the treatment. These will be categorised into adverse events likely related to the treatment and those likely unrelated. Serious adverse events are defined as any untoward and unexpected medical occurrence that results in death, is life-threatening, requires hospitalisation or prolongation of existing inpatient's hospitalisation, results in persistent or significant disability or incapacity, is a congenital anomaly or birth defect, or any other important medical condition which may require medical or surgical intervention to prevent one of the outcomes listed.

34. Statistical Software

Stata v16.1 will be used (StataCorp. Stata Statistical Software: Release 16.1. College Station, TX: StataCorp LLC)

35. References

1. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, Bridgett L, Williams S, Guillemin F, Hill CL *et al*: **The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study**. *Annals of the rheumatic diseases* 2014, **73**:1323-1330.
2. Murphy LB, Helmick CG, Schwartz TA, Renner JB, Tudor G, Koch GG, Dragomir AD, Kalsbeek WD, Luta G, Jordan JM: **One in four people may develop symptomatic hip osteoarthritis in his or her lifetime**. *Osteoarthritis Cartilage* 2010, **18**(11):1372-1379.
3. Australia A: **Counting the cost: Part 1: Healthcare costs - The current and future burden**. 2016.
4. Ackerman IN, Bohensky MA, de Steiger R, Brand CA, Eskelinen A, Fenstad AM, Furnes O, Graves SE, Haapakoski J, Makela K *et al*: **Lifetime Risk of Primary Total Hip Replacement Surgery for Osteoarthritis From 2003 to 2013: A Multinational Analysis Using National Registry Data**. *Arthritis Care Res (Hoboken)* 2017, **69**(11):1659-1667.
5. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, Kraus VB, Lohmander LS, Abbott JH, Bhandari M *et al*: **OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis**. *Osteoarthritis and Cartilage* 2019, **27**(11):1578-1589.
6. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, Callahan L, Copenhaver C, Dodge C, Felson D *et al*: **2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee**. *Arthritis Care Res (Hoboken)* 2020, **72**(2):149-162.
7. Royal Australian College of General Practitioners: **Guideline for the non-surgical management of hip and knee osteoarthritis**. In. Melbourne: RACGP; 2018.
8. Moseng T, Dagfinrud H, Smedslund G, Osteras N: **The importance of dose in land-based supervised exercise for people with hip osteoarthritis. A systematic review and meta-analysis**. *Osteoarthritis Cartilage* 2017, **25**(10):1563-1576.
9. Moseng T, Dagfinrud H, Smedslund G, Osteras N: **Corrigendum to 'The importance of dose in land-based supervised exercise for people with hip osteoarthritis. A systematic review and meta-analysis' [Osteoarthritis Cartilage 25 (2017) 1563-1576]**. *Osteoarthritis Cartilage* 2018, **26**(5):707-709.
10. Loureiro A, Mills PM, Barrett RS: **Muscle weakness in hip osteoarthritis: A systematic review**. *Arthritis Care Res* 2013, **65**:340-352.

11. Hall M, Wrigley TV, Kasza J, Dobson F, Pua YH, Metcalf BR, Bennell KL: **Cross-sectional association between muscle strength and self-reported physical function in 195 hip osteoarthritis patients**. *Semin Arthritis Rheum* 2017, **46**(4):387-394.
12. Wallis JA, Webster KE, Levinger P, Taylor NF: **What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis**. *Osteoarthritis and Cartilage* 2013, **21**(11):1648-1659.
13. old WHOIsgropafh-y: **Information sheet: global recommendations on physical activity for health 18-64 years old**. 2011.
14. Philbin EF, Groff GD, Ries MD, Miller TE: **Cardiovascular fitness and health in patients with end-stage osteoarthritis**. *Arthritis and rheumatism* 1995, **38**(6):799-805.
15. Jung JH, Seok H, Kim JH, Song GG, Choi SJ: **Association between osteoarthritis and mental health in a Korean population: a nationwide study**. *International journal of rheumatic diseases* 2018, **21**(3):611-619.
16. Reyes C, Leyland KM, Peat G, Cooper C, Arden NK, Prieto-Alhambra D: **Association Between Overweight and Obesity and Risk of Clinically Diagnosed Knee, Hip, and Hand Osteoarthritis: A Population-Based Cohort Study**. *Arthritis & rheumatology (Hoboken, NJ)* 2016, **68**(8):1869-1875.
17. Calders P, Van Ginckel A: **Presence of comorbidities and prognosis of clinical symptoms in knee and/or hip osteoarthritis: A systematic review and meta-analysis**. *Semin Arthritis Rheum* 2018, **47**(6):805-813.
18. Dunlop DD, Semanik P, Song J, Manheim LM, Shih V, Chang RW: **Risk factors for functional decline in older adults with arthritis**. *Arthritis and rheumatism* 2005, **52**(4):1274-1282.
19. White DK, Neogi T, Nguyen US, Niu J, Zhang Y: **Trajectories of functional decline in knee osteoarthritis: the Osteoarthritis Initiative**. *Rheumatology (Oxford, England)* 2016, **55**(5):801-808.
20. Bieler T, Siersma V, Magnusson SP, Kjaer M, Christensen HE, Beyer N: **In hip osteoarthritis, Nordic Walking is superior to strength training and home-based exercise for improving function**. *Scandinavian journal of medicine & science in sports* 2017, **27**(8):873-886.
21. Timmons JF, Minnock D, Hone M, Cogan KE, Murphy JC, Egan B: **Comparison of time-matched aerobic, resistance, or concurrent exercise training in older adults**. *Scandinavian journal of medicine & science in sports* 2018, **28**(11):2272-2283.
22. Villareal DT, Aguirre L, Gurney AB, Waters DL, Sinacore DR, Colombo E, Armamento-Villareal R, Qualls C: **Aerobic or Resistance Exercise, or Both, in Dieting Obese Older Adults**. *The New England journal of medicine* 2017, **376**(20):1943-1955.
23. Bellamy N, Carette S, Ford P, Kean W, le Riche N, Lussier A, Wells G, Campbell J: **Osteoarthritis antirheumatic drug trials. II. Tables for calculating sample size for clinical trials**. *The Journal of rheumatology* 1992, **19**(3):444-450.
24. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, Forbes A, Dobson F, Abbott JH, Dalwood A *et al*: **Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial**. *Pain* 2018, **159**(9):1833-1842.
25. Bennell KL, Egerton T, Martin J, Abbott JH, Metcalf B, McManus F, Sims K, Pua Y-H, Wrigley TV, Forbes A *et al*: **Effect of Physical Therapy on Pain and Function in Patients With Hip Osteoarthritis: A Randomized Clinical Trial**. *JAMA* 2014, **311**(19):1987-1997.
26. Wasserstein RL, Schirm AL, Lazar NA: **Moving to a World Beyond "p < 0.05"**. *The American Statistician* 2019, **73**(sup1):1-19.
27. GA B: **Psychophysical bases of perceived exertion.** *Med Sci Sports Exerc* 1982, **14**:377–381.
28. White IR, Horton NJ, Carpenter J, Pocock SJ: **Strategy for intention to treat analysis in randomised trials with missing outcome data**. *Bmj* 2011, **342**:d40.
29. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN: **The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research**. *Arthritis and rheumatism* 2003, **49**(2):156-163.
30. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ: **The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research**. *Psychiatry Res* 1989, **28**(2):193-213.

Appendix 1

VARIABLES IN THE DATA SET

Name	Description	Scale	Time-points measured	Variable name in dataset
Primary Outcomes				
Severity of overall hip pain	Scored on an 11-point numerical rating scale (NRS) for average overall hip pain in the last week.	Ranges from 0 to 10; where 0=no pain and 10=worst pain possible.	Baseline, 3 months and 9 months	nrs_pain_0w nrs_pain_3m nrs_pain_9m PHOENIX database
Physical function subscale of the Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC)	Scored using 17 questions regarding hip function in the last 48 hours with Likert response options ranging from None to Extreme.	Sum 17 questions: womac_c8 to womac_c24. Total score ranges from 0 to 68; higher scores indicate worse function.	Baseline, 3 months and 9 months	Derived womac_function0/3/9 PHOENIX database
Secondary Outcomes				
Overall hip pain during walking	Scored on an 11-point numerical rating scale (NRS) for overall hip pain during walking in the last week.	Ranges from 0 to 10; where 0=no pain and 10=worst pain possible.	Baseline, 3 months and 9 months	nrs_walking_0w nrs_walking_3m nrs_walking_9m PHOENIX database
Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC) Pain subscale	Scored using 5 questions regarding hip pain in the last 48 hours with Likert response options ranging from None to Extreme.	Sum 5 questions after recoding responses down by 1: womac_a_1 to womac_a_5. Ranges from 0 to 20; higher scores indicate worse pain	Baseline, 3 months and 9 months	Derived womac_pain 0/3/9 PHOENIX database
Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC) Stiffness subscale	Scored using 2 questions regarding hip stiffness in the last 48 hours with Likert response options ranging from None to Extreme.	Sum 2 questions after recoding responses down by 1: womac_b_6 to womac_b_7. Ranges from 0 to 8; higher scores indicate worse stiffness	Baseline, 3 months and 9 months	Derived womac_stiffness0/3/9 PHOENIX database
Global improvement in:	Scored using a 7-point global rating of	Participants who indicate that they are “moderately better”	3 months and 9 months	Derived- see Appendix 2.

a) Hip pain, b) Physical function	change Likert scale from with response options ranging from “much worse” to “much better” when compared to baseline.	or “much better” will be classified as improved. All other respondents will be classified as not improved.		globalpain__improved3 globalfunction_improved3 globalpain_improved 9 globalfunction_improved9 PHOENIX database
Quality of life (AQoL-6D)	Scored using the 20-item Assessment of Quality of Life II Instrument (6D version), which covers the topics of Independent Living, Relationships, Mental Health, Coping, Pain and Senses to come up with one overall value representing quality of life.	Total score ranges from −0.04 to 1.00; higher scores indicate better quality of life; calculated using https://www.aqol.com.au/index.php/scoring-algorithms, AQoL-6D utility algorithms, STATA-6D, accessed 3rd Oct, 2023.	Baseline, 3 months and 9 months	Derived AQOL _6D _0 AQOL _6D _3 AQOL _6D _9 PHOENIX database
*Hip extensor strength	Maximum voluntary isometric strength will be assessed using hand held dynamometer (Lafayette Manual Muscle Test System) with participants in supine and hip in a neutral position. Mean torque over 2 maximal efforts lasting 3 seconds will be recorded (Nm/kg)	Higher score indicates greater strength.	Baseline and 3 months	hip_ex_0w hip_ex_3m PHOENIX F2F Testing database
*Hip abductor strength	Maximum voluntary isometric strength will be assessed using hand held dynamometer (Lafayette Manual Muscle Test System) with participants in supine and hip in a neutral position. Mean torque over 2 maximal efforts lasting 3 seconds will be recorded (Nm/kg)	Higher score indicates greater strength.	Baseline and 3 months	hipab_0w hipab_3m PHOENIX F2F Testing database
*Knee extensor	Maximum voluntary	Higher score indicates greater	Baseline and 3	knee_ext_0w

strength	isometric strength will be assessed using an isokinetic dynamometer with the knee at 60 degree knee flexion. Peak torque over 3 maximal efforts lasting 5 seconds will be recorded (Nm/kg)	strength.	months	knee_ext_3 m PHOENIX F2F Testing database
Physical function: Patient specific functional scale	Up to five activities of daily living are identified and scored on an 11-point NRS to record if performance was limited in the previous week.	Ranges from 0 to 10; where 0=unable to perform and 10=able to perform the activity at the same level as before hip osteoarthritis. Higher scores indicate better function.	Baseline, 3 months and 9 months	Derived psfs0/3/9 PHOENIX database
*30 second chair sit to stand test	Number of complete chair stands completed in 30 secs	Higher score indicates better function.	Baseline and 3 months	sts_0w sts_3m PHOENIX F2F Testing database
*6-step stair climb test	Time take to ascend and descend a flight of 6 stairs as quickly as possible. Use of handrail is permitted if needed.	Shorter time indicates better function.	Baseline and 3 months	stair_time_0 w stair_time_3 m PHOENIX F2F Testing database
*40 meter fast pace walk test	Time to taken to walk 4 x 10 m quickly but safely (m/s)	Higher speed indicates better function.	Baseline and 3 months	total_40m_ti me_0w total_40m_ti me_3m PHOENIX F2F Testing database
*Cardiorespiratory fitness	Prior to this assessment hydration is assessed using a bioelectrical impedance scan. However, if participants have a	Oxygen consumption in ml per min relative to body mass will be reported. Higher values indicate greater cardiorespiratory fitness.	Baseline and 3 months	VO2_ml_kg_ min_0/3 PHOENIX VO2 database

	non-titanium metal implant they will be asked to provide a urine sample prior to cardiorespiratory assessment instead. Cardiorespiratory fitness will be assessed as submaximal oxygen consumption using a graded incremental exercise protocol using a cycle ergometer. Participants will start at 1W/kg and increase 0.5W/body mass (kg). Test will terminate when respiratory quotient is >1.00, or equal to ~75% of max heart rate.			
Other measures				
Adherence to consultation with physiotherapist	Recorded by physiotherapist for each consultation	Number of consultations of a maximum of 9 attended (face-to-face or telephone)	Throughout the initial 3 months	Counts of 1 from Consultation_1 to Consultation_9 will be summed PHOENIX_Adherence
Number of strength sessions at least “very hard”(RPE)	Obtained from participant log book (Range from 0 to 36 sessions). Reported in whole numbers. Reported number per week and overall (average over 12 weeks).	Higher indicates more strengthening exercise completed	Throughout the initial 3 months	Calculate average RPE for each day (session) across 3-6 exercises completed (1 day = 1 strength session) for each day/wk. Count as 1 and sum for each week and overall 12 weeks if observation for day*_av_wk * is 7-8 RPE (= very hard)

				or higher (9-10) and not missing PHOENIX_Attendance_Fidelity
Total duration in minutes of aerobic physical activity working at least somewhat hard	Obtained from participant log book each week up to week 12.	Higher indicates more exercise completed	Throughout the initial 3 months	Aerobic_mins_somewhat hard derived from sum of days and weeks PHOENIX Adherence
Number of activity minutes at minimum prescribed heart rate	Participants will be instructed to wear a wearable physical activity monitor over the initial 3 months of the study period. Number of minutes of at least moderate intensity exercise will be based on heart rate data. With participant consent, de-identified daily heart rate data will be automatically uploaded to a cloud-based system (Fitabase – Small Step Lab) from the participant's wearable activity monitor account.	Higher indicates greater time being active	Throughout the initial 3 months	Week1-Week13 PHOENIX Garmin Heart Rate Data
Treatment fidelity	We will use physiotherapists' treatment notes to determine whether in at least 7 of the 9 consultations, the following criteria were fulfilled: i) 3 strengthening exercises (from the first consultation) and 4–6 strengthening exercises (from the second consultation onwards) of at least "very hard" intensity were prescribed, and for participants in the	Number of exercise prescriptions according to protocol; binary variable of whether criteria were fulfilled will be derived as well as presenting average number of strength and average minutes of aerobic exercises per week prescribed as per criteria.	Throughout the initial 3 months	PHOENIX Attendance Fidelity

	combined aerobic physical activity and strengthening group: ii) prescribed an increase in moderate intensity aerobic physical activity by at least 10 min per week, until 150 min per week was reached. Both criteria must be satisfied for participants in the combined aerobic physical activity and strengthening group, while only the strength criteria is applicable to the strengthening only group.			
Adverse events	Recorded by participants using a custom-developed table.	Number and nature (related, non-related, severity) of adverse events will be reported.	3 months and 9 months	Derived-Count each Code (1-10 at 3 months; 1-7 at 9 months) PHOENIX_Adverse Events database
Co-intervention use	Participants will complete a custom-developed table to indicate the frequency of use (over the past 6 months) of a range of pain and arthritis medications and co-interventions.	Participants who indicate they have used a drug/supplement at least once a week in the past month will be reported as a current user of the relevant medication. Participants who have used a co-intervention once in the past 6 months will be reported as a current user.	Baseline, 3 months and 9 months	co_in1-14; meds_final1-5 (derived, see Appendix 2) PHOENIX database
Overall satisfaction with the exercise program	Scored on a Likert Scale for "Overall how satisfied are you with the program you received for your hip pain? From 0 = 'not at all satisfied' to 10 = 'extremely satisfied'.	Higher indicates greater satisfaction	6 weeks and 3 months	satisfaction_6wk satisfaction_3m PHOENIX database
Therapeutic alliance using the Working Alliance Inventory Short Form	Scored separately by the participant after 6th consultation, based on 12 statements relating to the perceived trust and agreement	Overall scores range from 12 to 84 (with higher scores indicating a stronger therapeutic alliance)	6 weeks	work_all_6wk PHOENIX database

	between the therapist and client, using a 7-point scale.			
Baseline measures				
Physiotherapist	Physiotherapist participant randomized to after baseline assessment complete	1,2,3,4,5,6,7,8.	Randomisation	Physiotherapist PHOENIX database
Group	Group participant randomized to after baseline assessment complete	Aerobic+ strengthening vs strengthening group.	Randomisation	Group PHOENIX_group allocation database
Age	Calculated from date of birth (date_of_birth) to date randomised (date_randomised). Dates reported in MM/DD/YYYY	Reported in years, no rounding.	Baseline	Derived- see Appendix 2. age = date_randomised minus date_of_birth PHOENIX database
Gender	Self-report of identified gender	Male or female or not disclosed	Baseline	gender_0w PHOENIX database
*Height	Measured in the laboratory using a stadiometer	Measured in metres	Baseline	height_0w PHOENIX database
*Weight	Measured in the laboratory using scales	Measured in kilograms	Baseline	weight_0w PHOENIX database
Body mass index (BMI)	Calculated from height and weight.	Reported in kg/m^2	Baseline	Derived- see Appendix 2. bmi_0w PHOENIX database
Current employment status	Participants will report their current employment status using a categorical scale with response option currently employed; retired (not due to health reasons); unemployed/student; homemaker; unable to work due to health reasons.	The number and proportion of respondents for each category will be reported.	Baseline	employment_0w PHOENIX database

Education level	Participants will report their current education level using a categorical scale with response options < 3 years of high school; 3 or more years of high school; some education beyond high school; completed tertiary or higher education	The number and proportion of respondents for each category will be reported.	Baseline	education_0w PHOENIX database
Unilateral hip symptoms	Participant will self-report if their symptoms are in one or both hips.	Reported as number and percentage	Baseline	bilat_unilat_0w PHOENIX database
Duration of hip symptoms in the study hip	Participants will self-report the total duration of time since their study hip symptoms began.	Reported in years	Baseline	duration_symptoms_0w PHOENIX database
Co-morbid conditions	Self-administered comorbidity questionnaire[29]	The number and proportion of comorbidities will be reported.	Baseline	heart_0w hbp_0w dep_0w anemia_0w ulcer_0w ra_0w diab_0w osteo_0w lung_0w cancer_0w back_0w kidney_0w liver_0w other_prob_1_0w other_prob_2_0w in Baseline Demographic Data sheet in PHOENIX database
*Radiographic disease severity of hip osteoarthritis	Rated from a non-weight bearing x-ray using the KL scale (0,1,2,3,4).	The number and proportion of participants with Grade 0 and Grade 1 versus Grade 2 (mild disease severity), 3 (moderate disease) and Grade 4 (severe disease) will be reported.	Baseline	See derived variable, Appendix 2. KL_Grade_0w PHOENIX F2F Testing

				database
Expectation of treatment outcome	Rated using a Likert scale with anchors of “no effect at all” to “complete recovery”	The number and proportion of participants selecting each response option will be reported.	Baseline	treatexp_0w PHOENIX database
Pain type	The painDETECT questionnaire is used to determine how likely pain has a neuropathic component. The tool includes questions about hip pain intensity (NRS, 0 to 10 where higher scores indicate greater severity), pain course pattern (option from four illustrations), pain radiation (yes/no question) and somatosensory phenomena (seven questions on a 6-point Likert Scale, 0 to 5 with higher scores indicating likelihood of somatosensory phenomena).	Total scores, derived by adding pain course pattern, pain radiation and each of the seven somatosensory phenomena questions, range from –1 to 38 with higher scores indicating more neuropathic-like symptoms. Scores of <= 12 indicate pain is unlikely to be neuropathic and scores of > 19 suggest pain is likely to have neuropathic component.	Baseline	PainDETECT_score_0w derived from adding course_0w radiating_pain_0w burning_0w tingling_0w light_touch_0w sudden_attacks_0w cold_heat_0w numbness_0w light_pressure_0w PHOENIX PainDETECT database
Sleep quality	The Pittsburgh Sleep Quality Index (PSQI) questionnaire is used to evaluate sleep quality and fatigue, respectively over the past month. The PSQI measures seven components of sleep quality.	Each component is scored from 0 to 3 creating a total score from 0 to 21, with higher scores indicating poorer sleep quality[30].	Baseline	Derived global_PSQI_Score_0w from summing 7 components, each component derived from Pittsburgh PSQI case report form PHOENIX database

Please note:

*These secondary outcomes were only collected if face-to-face assessments were permitted and deemed safe by the University of Melbourne and the Victorian Government (COVID-19).

Appendix 2

DEFINITIONS OF DERIVED VARIABLES IN THE DATA SET

Variable in data set	Unit	Calculation	Range	Better	Variable label in spreadsheet
Age	Reported in years, no rounding.	Calculated from date of birth (date_of_birth) to date randomised (date_randomised).	>=45 years	NA	age PHOENIX database
Body mass index	kg/m²	Body mass in kilograms ("Weight") divided by the square of "Height" (in metres).	N/A	N/A	bmi_0w PHOENIX database
Change in severity of overall average hip pain in the past week (primary outcome):	11-point numeric rating scale.	Change score at 3 months (primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.	-6 (inclusion criteria of >=4, so if rate 4 at baseline and 10 at follow-up then 4-10 = pain worse by 6 units) to 10.	Higher is better; positive means improved from baseline.	nrs_pain_0w nrs_pain_3m nrs_pain_9m PHOENIX database
Change in difficulty with	Western Ontario and McMaster	Change score at 3 months	-68 to 68	Higher is better;	WOMAC_C_0w

physical function (primary outcome):	Universities Osteoarthrits Index (Likert version 3.1) score.	(primary time point) and 9 months (secondary time point) will be calculated as baseline minus follow-up.		positive means improved from baseline.	WOMAC_C_3m WOMAC_C_9m PHOENIX database
Change in severity of hip pain while walking:	Numeric rating scale.	Change score at 3 months and 9 months will be calculated as baseline minus follow-up.	-10 to 10	Higher is better; positive means improved from baseline.	nrs_walking_0w nrs_walking_3m nrs_walking_9m PHOENIX database
Change in each of WOMAC i) pain and ii) stiffness in the last week:	Using subscales of the Western Ontario and McMaster Universities Osteoarthrits Index (Likert version 3.1), pain (5-items); ii) stiffness (2-items) will be self-assessed.	Scores will be calculated for pain and stiffness and range from 0 to 20, and 0 to 8 respectively with higher scores indicating greater symptom severity. Change score at 3 months and 9 months will be	Change in pain: -20 to 20; change in stiffness: -8 to 8.	Higher is better; positive means improved from baseline.	I) WOMAC_A_0w WOMAC_A_3m WOMAC_A_9m ii) WOMAC_B_0w WOMAC_B_3m WOMAC_B_9m PHOENIX database

		calculated as baseline minus follow-up.			
Change in health-related quality of life:	The AQoL questionnaire (version AQoL-6D) measures health-related quality of life.	This is a 20-item questionnaire and scores range from -0.04 to 1.00 with 1.00 indicating full health-related quality of life. Change score at 3 months and 9 months will be calculated as baseline minus follow-up.	-1.04 to 1.04.	Lower is better; negative means improved from baseline.	AQOL_0w AQOL_3m AQOL_9m PHOENIX database
Change in muscle strength:	Maximal normalised isometric hip abduction, hip extension and knee extensor strength (peak torque, Nm/kg) will be assessed. The sum of hip abduction, hip extensor and knee extensor strength is calculated and will be reported as an overall strength score.	Change score at 3 months will be calculated as baseline minus follow-up.	Unlimited	Lower is better; negative values indicate improvement from baseline.	Dmuscle_strength3 PHOENIX F2F Testing database
Change in	At baseline,	Change score at	-10 to 10	Lower is	PSFS_0w

patient specific function scale:	participants indicate up to five activities of daily living in which performance was limited in the previous week and rate these on an 11-point NRS ranging from 0-10, where 0 indicates 'unable to perform' and 10 indicates 'able to perform at the same level as before injury or problem'. Total score is the sum of the ratings for each activity divided by the number of activities, with higher scores indicating better function [39].	3 months and 9 months will be calculated as baseline minus follow-up.		better; negative means improved from baseline.	PSFS_3m PSFS_9m PHOENIX database
Change in physical performance using 3 tests	Physical performance is measured using the Osteoarthritis Research Society International set of performance-based measures of physical function, including the i) 30-second sit-to-stand test	Change scores at 3 months will be calculated as baseline minus follow-up.	Unlimited	Lower is better for 30s sit-to stand test, otherwise higher is better; negative values indicate improveme	i) sts_0w sts_3m ii) walk40m_0w walk40m_3m iii) stair_0w stair_3m PHOENIX F2F

	(number of sit-to-stands completed), ii) 40-meter fast pace walk test (time) and iii) 6-step stair climb test (time).			nt from baseline for 30 s sit-to-stand test, otherwise positive values indicate improveme nt.	Testing database
Change in cardiorespiratory fitness:	At baseline and 3 months, a submaximal exercise test is performed on a cycle ergometer, and submaximal cardiorespiratory fitness is measured by breath-by-breath indirect calorimetry using a Vmax Encore metabolic cart. Cardiorespirator fitness will be expressed as the VO_2 in milliliters per kilogram at which the respiratory exchange ratio of 1.000.	Change scores at 3 months (primary time point) will be calculated as baseline minus follow-up.	Unlimited	Lower is better; negative scores indicate improveme nt from baseline.	w_VO2_ml_kg_min m_VO2_ml_kg_min PHOENIX VO2 database
Global improvement (perceived	Scored using a 7-point global rating of change: Likert	Participants who indicate that they are	N/A	N/A	global_pain_3m global_function

change) in: a) Hip pain, b) Physical function	scale from with response options ranging from “much worse=1” to “much better=7” at 3 months and at 9 months when compared to baseline.	“moderately better=6” or “much better=7” will be classified as improved. All other respondents will be classified as not improved.			_9m global_pain_3m global_function_9m PHOENIX database
Full adherence	0,1.	Full adherence to the aerobic component for the aerobic and strengthening group=1: Report completing the minimum number of minutes at at least a somewhat hard RPE set by physiotherapist for 8 weeks of the 3 months.	N/A	N/A	Full_Adh
Complete primary outcomes	0,1.	Those participants who provide both primary outcomes at 3 months will be coded as 1; and	N/A	N/A	complete

		those participants who are missing one or both primary outcomes at 3 months will be coded as 0.			
Radiographic disease severity of hip osteoarthritis as a binary variable (KL grade =<2 versus KL grade > 2)	0 = KL grade =<2; 1 = KL grade > 2	KL Grade 0-1 versus Grade 2-4.	N/A	N/A	KLgrade2plus derived from KL_Grade_0w PHOENIX F2F Testing database
Satisfactory adherence to strengthening exercises	Derive as >= 3 strength sessions minimum per week for at least 10 weeks at at least very hard RPE = 1, 0 otherwise.	Derive from number of strength sessions at least very hard	NA	NA	satisfactory_str_adh PHOENIX Adherence
Self-report completing number of minutes at the minimum heart rate set by the physiotherapist, aerobic group only	Derived from participant logbook: total duration in minutes of aerobic physical activity working at least somewhat hard during each week up to week 12.	Derive "completed number of minutes" =1 if completed minutes >= prescribed minutes	NA	NA	PHOENIX Adherence
Satisfactory adherence to the strengthening component or the combined aerobic and strengthening	0,1	Satisfactory adherence as defined in Section 19 =1; 0 otherwise.	N/A	N/A	Satisfactory_aer_adh

component, as relevant					
At least 1 comorbidity	0,1	At least one of heart_0w hbp_0w dep_0w anemia_0w ulcer_0w ra_0w diab_0w osteo_0w lung_0w cancer_0w back_0w kidney_0w liver_0w other_prob_0w	NA	NA	Atleastonecomobidity
At least 1 treatment at baseline, 3months and 9 months (each timepoint separately)	0,1	At least one of co_in1_0w co_in2_0w co_in3_0w co_in4_0w co_in5_0w co_in6_0w co_in7_0w co_in8_0w co_in9_0w co_in10_0w co_in11_0w co_in12_0w co_in13_0w co_in14_0w. Same for timepoints 3months and 9 months.	NA	NA	Atleastonetreatment_0w Atleastonetreatment_3m Atleastonetreatment_9m

At least 1 medication at baseline, 3months and 9 months (each timepoint separately)	0,1	meds_final1_0w = Non-steroidal anti-inflammatory drugs (meds1_0w =Anti-inflammatory tablets or capsules (eg. Voltaren, Nurofen, Mobic) & meds2_0w =Anti-inflammatory tablets or capsules - cox-2 inhibitors (eg. Celebrex)) meds_final2_0w =Acetaminophen (meds3_0w = Paracetamol only (e.g. Panadol) & meds4_0w Paracetamol plus codeine combinations (e.g. Panadeine forte)) meds_final3_0w =Topical anti-inflammatory	NA	NA	meds_final1_0w meds_final2_0w meds_final3_0w meds_final4_0w meds_final5_0w Same for 3 month and 9-month timepoints.

		drugs (meds6_0w =Topical anti-inflammatory gels or creams (e.g. Voltaren emugel)) meds_final4_0w =Oral corticosteroids (meds8_0w =Oral Corticosteroids (e.g. prednisolone, dexmethsone)) meds_final5_0w =Oral opioids (meds5_0w = Tramadol & meds10_0w =Opioid Oral Medication (e.g. oxycodeine, morphine, ms-contin, oxycontin, kapanol))			
At least 1 medication	0,1	At least one of meds_final1_0w meds_final2_0w meds_final3_0w meds_final4_0w meds_final5_0w	NA	NA	Atleastonemedication_0w Same for 3 month and 9-month timepoints.

		Same for 3 month and 9-month timepoints.			

Appendix 3

LIST OF TABLES/FIGURES/LISTINGS

Number	Title – analysis set
Table 1	Baseline characteristics of participants by group, reported as mean (standard deviation) unless otherwise stated.
Table 2	Outcome measures over time across treatment groups for continuous outcomes.
Table 3	Change in continuous outcome measures within groups and between groups over time using complete case/multiply-imputed data.
Table 4	Binary secondary outcomes and adjusted relative risks and risk differences using complete case/multiply imputed data.
Table 5	Potential binary moderator of the effect of the addition of aerobic exercise to a strengthening program on primary outcomes at 3 months using complete case data.
Table 6	Potential continuous moderator of the effect of the addition of aerobic exercise to a strengthening program on primary outcomes at 3 months using complete case data.
Table 7	Effects of the combined aerobic and strengthening intervention on the primary outcomes at 3-months assuming full adherence.
Table 8	Process measures.
Table 9	Adherence measures.
Table 10	Adverse events, pain medication use and other co-interventions.
Appendix Table 1	Baseline characteristics of participants who provided both primary outcomes at 3-months and those participants who did not.
	Appendix tables using complete case data if use multiply imputed data as primary analysis (Table 3 and 4).

Appendix 4

TABLE TEMPLATES

Table 1. Baseline characteristics of participants by group, reported as mean (standard

deviation, SD) unless otherwise stated.

Baseline characteristic	Group 1 (n=xxx)	Group 2 (n=xxx)
Age, (years)		
Female, n (%)		
Height, (m)		
Body mass, (kg)		
Body mass index, (kg/m^2)		
Education level, n (%)		
Less than 3 years of high school		
3 or more years of high school		
Some education beyond high school		
Completed tertiary or higher education		
Current employment status, n (%)		
Currently employed		
Unable to work due to health reasons		
Retired (not due to health reasons)		
Unemployed/student		
Homemaker		
Hip symptom duration, (years)		
Unilateral symptoms, n (%)		
Pain type (PainDETECT) #		
Comorbid conditions, n (%)		
≥1 comorbid condition		
Heart disease		
High blood pressure		
Depression		
Anemia or other blood disease		
Ulcer or stomach disease		
Rheumatoid arthritis		
Diabetes		
Osteoarthritis		
Lung disease		
Cancer		
Back pain		
Kidney disease		
Liver disease		
Other		
Radiographic disease severity^, n (%)		
Grade <2		
Grade ≥2		
Expectation of treatment outcome, n (%) ‡		

No effect at all
Minimal improvement
Moderate improvement
Large improvement
Complete recovery
Treatments for hip in last 6 months, n (%)
≥1 treatment
Massage
Manual therapy
Gait aid
Ultrasound
Low level laser therapy (LLLT)
Transcutaneous Electrical Nerve Stimulation (TENS)
Hip braces
Heat/cold treatment
Land-based exercises (eg. strengthening, aerobic or stretching exercises)
Injections (eg. cortisone, Synvisc, platelet-rich plasma)
Hydrotherapy (eg warm water exercises)
Acupuncture
Arthroscopic surgery
Hip joint replacement surgery
Current pain medication use, n (%) *
≥1 medication used
Non-steroidal anti-inflammatory drugs
Acetaminophen
Topical anti-inflammatory drugs
Oral corticosteroids
Oral opioids
Pittsburgh Sleep Quality Index (PSQI)

SD=standard deviation; kg=kilograms; m=metres; IQR= interquartile range (25th to 75th percentile).

The painDETECT questionnaire includes questions about pain course pattern, pain radiation and somatosensory phenomena. Total scores range from –1 to 38 with higher scores indicating more neuropathic-like symptoms.

^N = 127 participants (n = xx in Group 1; n = xx in Group 2). 69 missing due to COVID-19 lockdowns and restrictions in Australia from March 2020.

‡ Scored from a question asking about expectation of study treatment outcomes, with self-reported scores on a

5-point Likert scale from “0=no effect at all” to “4=complete recovery”.

* Defined as ≥ 1 time per week over the last 6 months.

Table 2. Outcome measures over time across treatment groups for continuous outcomes.

	Baseline		3-months		9-months	
	Group 1	**Group 2**	**Group 1**	**Group 2**	**Group 1**	**Group 2**
	N=xxx	**N=xxx**	**N=xxx**	**N=xxx**	**N=xxx**	**N=xxx**
Primary outcomes						
Overall hip pain, past week (NRS) ‖						
Physical function, past 48 h (WOMAC) ¶						
Secondary outcomes						
Hip pain during walking (NRS) ‖						
Hip pain (WOMAC) ¶						
Hip stiffness (WOMAC) ¶						
Quality of life (AQoL-6D) †						
Physical function (PSFS) *						
Muscle strength (Nm/kg)					–	–
30 s sit to stand (number)					–	–
Timed stair climb (m/s)					–	–
40 m fast walk (seconds)					–	–
Submaximal cardiorespiratory fitness (ml/min)					–	–

NRS=numerical rating scale; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index; AQoL-6D=Assessment of Quality-of-Life instrument-6 dimension;

PSFS= Patient Specific Functional Scale.

Data presented are complete cases and summarised as mean (standard deviation).

|| NRS, 0–10 with higher scores indicating more pain; Minimal clinically important difference (MCID)=1.8.

¶ WOMAC, 0–68 for physical function (MCID=6), 0-20 for pain, 0–8 for stiffness with higher scores indicating more dysfunction/pain/stiffness.

† AQoL-6D, -0.04–1.00 with higher scores indicating better quality of life.

* Physical function (PSFS), 0–10 with higher scores indicating better ability to perform self-selected activities at the same level as before hip problem.

Table 3. Change in continuous outcome measures within groups and between groups over time using complete case/multiply imputed data.

	Mean (SD) change within groups Baseline minus 3-months		Difference in change between groups at 3-months Group 1 vs Group 2		Mean (SD) change within groups Baseline minus 9-months		Difference in change between groups at 9-months Group 1 vs Group 2	
	Group 1 (N=xxx)	Group 2 (N=xxx)	Mean (95% CI) (N=xxx)	P-value	Group 1 (N=xxx)	Group 2 (N=xxx)	Mean (95% CI) (N=xxx)	P-value
Primary outcomes								
Overall hip pain, past week (NRS) †								
Physical function, past 48 h (WOMAC) †								
Secondary outcomes								
Hip pain during walking (NRS) †								
Hip pain (WOMAC) †								
Hip stiffness (WOMAC) †								
Quality of life (AQoL-6D) *								
Physical function (PSFS) *								
Muscle strength (Nm/kg) *					–	–	–	–
30 s sit to stand (number) †					–	–	–	–
Timed stair climb (m/s) *					–	–	–	–
40 m fast walk (seconds) *					–	–	–	–
Submaximal cardiorespiratory fitness (ml/min) *					–	–	–	–

SD=standard deviation; CI=confidence interval; NRS=numerical rating scale, 0–10 with higher scores indicating more pain, minimal clinically important difference (MCID)=1.8; WOMAC= Western Ontario and McMaster Universities Osteoarthritis Index, 0–68 for physical function (MCID=6), 0–20 for pain, 0–8 for stiffness with higher scores indicating more dysfunction/pain/stiffness; AQoL-6D=Assessment of Quality-of-Life instrument-6 dimension, -0.04–1.00 with higher scores indicating better quality of life; PSFS= Patient Specific Functional Scale, 0–10 with higher scores indicating better ability to perform self-selected activities at the same level as before hip problem.

† For change within groups, positive changes indicate improvement. For difference in change between groups, positive differences favor Group 1.

* For change within groups, negative changes indicate improvement. For difference in change between groups, negative differences favor Group 1.

Table 4: Binary secondary outcomes and adjusted relative risks and risk differences using complete case/multiply imputed data.

	Group 1 n/Total (%)	Group 2 n/Total (%)	Relative risk* (95% CI)	P-value	Risk Difference* (95% CI)	P-value
3-months						
Global improvement †						
In pain						
In function						
9-months						
Global improvement †						
In pain						
In function						

CI=confidence intervals.

* Relative risks < 1 and risk differences < 0 favor Group 1.

† Perceived change since baseline. Rated using a 7-point scale with terminal descriptors ‘much worse’ to ‘much better’, those indicating ‘moderately better’ or ‘much better’ classified as improved.

Table 5: Potential binary moderator of the effect of the addition of aerobic exercise to a strengthening program on primary outcomes at 3 months using complete case data.

Outcome (baseline minus 3-months)	Moderator	Group 1 Mean (SD)# [N=xxx]	Group 2 Mean (SD)# [N=xxx]	Group 2 – Group 1 Mean (95% CI) [N=xxx]	Interaction p-value
Overall hip pain, past week (NRS) †	Radiographic disease severity				
	Grade <2				
	Grade ≥2				
Physical function, past 48 h (WOMAC) ‡	Radiographic disease severity				
	Grade <2				
	Grade ≥2				

SD=standard deviation; CI=confidence intervals; NRS=numeric rating scale; WOMAC= Western Ontario and McMaster Universities Osteoarthritis Index.

#The effect on change in overall hip pain or change in physical function at 3-months of a radiographic disease severity grade of 0-1 or grade 2-4 at baseline in each of the intervention groups.

† NRS, 0-10 with higher scores indicating more pain; Minimal clinically important difference (MCID)=1.8.

‡ WOMAC, 0-68 with higher scores indicating more physical dysfunction, MCID=6.

Table 6: Potential continuous moderators of the combined aerobic and strengthening program on primary outcomes at 3 months using complete case data.

Outcome (baseline minus 3-months)	Moderator	Moderator coefficient# for Group 1		Moderator coefficient# for Group 2		Interaction between moderator and group	
		Mean (95% CI) [N=xxx]	P-value	Mean (95% CI) [N=xxx]	P-value	Difference in coefficients Group 2 – Group 1 (95% CI)	Interaction p-value
Overall hip pain, past week (NRS) *	BMI (kg/m^2)						
Physical function, past 48 h (WOMAC) †	BMI (kg/m^2)						
Overall hip pain, past week (NRS) *	Pain type (PainDETECT) ^						
Physical function, past 48 h (WOMAC) †	Pain type (PainDETECT) ^						
Overall hip pain, past week (NRS) *	Pittsburgh Sleep Quality Index (PSQI)						
Physical function, past 48 h (WOMAC) †	Pittsburgh Sleep Quality Index (PSQI)						

NRS=numeric rating scale; BMI=body mass index; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index.

Results presented in terms of a 1-unit increase in the potential continuous moderator in each group.

#Moderator coefficient: the effect on change in each primary outcome at 3-months of a 1-unit increase in the potential moderators in each of Group 1 and Group 2 groups.

* NRS, 0–10 with higher scores indicating more pain; Minimal clinically important difference (MCID)=1.8.

† WOMAC, 0–68 with higher scores indicating more physical dysfunction, MCID=6.

^ The painDETECT questionnaire includes questions about pain course pattern, pain radiation and somatosensory phenomena. Total scores range from –1 to 38 with higher scores indicating more neuropathic-like symptoms.

Table 7: Effects of the combined aerobic and strengthening intervention on the primary outcomes at 3-months assuming full adherence^.

	Difference in change between groups * Aerobic+strength – Strength	
	Mean (95% CI)	**P-value**
Overall hip pain, past week (NRS) †		
Physical function, past 48 h (WOMAC) ‡		

CI=confidence interval; NRS=numeric rating scale; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index; MCID=Minimal clinically important difference.

^ Full adherence defined as completing the number of minutes working at least somewhat hard as set by the physiotherapist in at least 8 weeks of the 3-month Aerobic+strength program and assumes adherence to strengthening exercises was equal between groups.

* For difference in change between groups, positive differences favor Aerobic+strength.

† NRS, 0–10 with higher scores indicate more pain; MCID=1.8; N=xxx.

‡ WOMAC, 0–68 with higher scores indicating more physical dysfunction, MCID=6; N=xxx.

Table 8: Process measures

Measure	Aerobic+Strength [n=xx]	Strength [n=xx]
Number of consultations attended		
Treatment fidelity #, n (%)		
Strengthening exercise ^		
Aerobic exercise ^^		-
Satisfaction with the intervention (0-10) *		
6-weeks		
3-months		
Therapeutic alliance ~		

Data presented as mean (standard deviation), unless otherwise stated.

Physiotherapists' treatment notes used to determine whether in at least 7 of the 9 consultations, the following criteria were fulfilled: i) 3 strengthening exercises (from the first consultation) and 4–6 strengthening exercises (from the second consultation onwards) of at least "very hard" intensity were prescribed, and for participants in the combined aerobic physical activity and strengthening group: ii) prescribed an increase in moderate intensity aerobic physical activity by at least 10 min per week, until 150 min per week was reached. Both criteria must be satisfied for participants in the combined aerobic physical activity and strengthening group, while only the strength criteria is applicable to the strengthening only group.

^ Number of strengthening exercise prescriptions according to intensity protocol - defined as 3 strengthening exercises from the first consultation and 4–6 strengthening exercises (from the second consultation onwards) of at least "very hard" intensity prescribed.

^^Number of exercise prescriptions that included an increase in moderate intensity aerobic physical activity by at least 10 min per week, until 150 min per week was reached

* Scored on a Likert Scale for "Overall how satisfied are you with the program you received for your hip pain? from 0 = 'not at all satisfied' to 10 = 'extremely satisfied'.

~ measured using the Working Alliance Inventory Short Form, scored by the participant after 6th consultation, based on 12 statements relating to the perceived trust and agreement between the therapist and client, each statement using 7-point scales. Overall scores range from 12 to 84 (with higher scores indicating a stronger therapeutic alliance).

Table 9: Adherence measures.

Measure	Group 1	Group 2
Number of consultations with physiotherapist, mean (SD):	xx (xx), N=xx	xx (xx), N=xx
Number of strength sessions at least "very hard" (RPE), mean (SD):		
Week 1		
Week 2		
Week 3		
Week 4		
Week 5		
Week 6		
Week 7		
Week 8		
Week 9		
Week 10		
Week 11		
Week 12		
Overall, mean (SD)		
Number of activity minutes at least "somewhat hard" (RPE), mean (SD):		
Week 1	-	
Week 2	-	
Week 3	-	
Week 4	-	
Week 5	-	
Week 6	-	
Week 7	-	
Week 8	-	
Week 9	-	
Week 10	-	
Week 11	-	
Week 12	-	
Overall, mean (SD)	-	
Number (%) classified satisfactorily adherent at 3-months *		
Number (%) classified fully adherent at 3-months †	-	

SD: standard deviation; RPE: rating of perceived exertion.

* Participant exercise logbooks are used to determine if 3 strengthening exercises (from the first consultation) and 4–6 strengthening exercises (from the second consultation onwards) at least at a "very hard" intensity (Rating of Perceived Exertion, RPE) were performed, at least three times per week. Adherence for the strengthening group is considered satisfactory if participants self-report completing strengthening exercises "very hard" at least 10 weeks of the 3 months. Adherence for the combined aerobic physical activity and strengthening exercise group, is considered satisfactory if participants self-report completing strengthening exercises "very hard" and self-report completing the number of minutes working at least "somewhat hard" (RPE) as set by the physiotherapist in at least 10 weeks of the 3-month program.

† Fully adherent defined as completing the number of minutes working at least "somewhat hard" (RPE) as set by the physiotherapist in at least 8 weeks of the 3-month program.

Table 10: Adverse events, pain medication use and other co-interventions.

Characteristic	Group 1	Group 2
Adverse events, 0 to 3-months *		
Discontinued due to related adverse event		
Any serious adverse events †		
Non-serious related adverse events ^		
Hip pain		
Knee pain		
Back pain		
Lower limb pain		
Foot pain		
Neck pain		
Hernia		
Skin irritation (Eg., caused by Garmin device)		
Dizziness/Faint		
Other		
Adverse events, 3 to 9-months *		
Discontinued due to related adverse event		
Any serious adverse events †		
Non-serious related adverse events ^		
Hip pain		
Knee pain		
Back pain		
Lower limb pain		
Foot pain		
Neck pain		
Other		
Pain medication use, 0 to 3-months ‡		
≥1 medication		
Acetaminophen		
Topical anti-inflammatory drugs		
Non-steroidal anti-inflammatory drugs		
Oral corticosteroids		
Oral opioids		
Pain medication use, 3 to 9-months ‡		
≥1 medication		
Acetaminophen		
Topical anti-inflammatory drugs		
Non-steroidal anti-inflammatory drugs		

Oral corticosteroids

Oral opioids

Other co-interventions, 0 to 3-months §

≥1 treatment

Massage

Manual therapy

Gait aid

Injections (eg. cortisone, Synvisc, platelet-rich plasma)

Ultrasound

Hip brace

Land-based exercises (eg. strengthening, aerobic or stretching exercises)

Low level laser therapy (LLLT)

Acupuncture

Transcutaneous Electrical Nerve Stimulation (TENS)

Heat/cold treatment

Hydrotherapy (eg warm water exercises)

Arthroscopic surgery

Hip joint replacement surgery

Other co-interventions, 3 to 9-months §

≥1 treatment

Massage

Manual therapy

Gait aid

Injections (eg. cortisone, Synvisc, platelet-rich plasma)

Ultrasound

Hip brace

Land-based exercises (eg. strengthening, aerobic or stretching exercises)

Low level laser therapy (LLLT)

Acupuncture

Transcutaneous Electrical Nerve Stimulation (TENS)

Heat/cold treatment

Hydrotherapy (eg warm water exercises)

Arthroscopic surgery

Hip joint replacement surgery

Data expressed as number of participants/total number of participants who provided data (percentage).

* Adverse events defined as any problem experienced as a result of the study, in the study hip or elsewhere in the

body, lasting for ≥2 days, and/or requiring cessation of the study intervention and/or seeking treatment from a healthcare professional. Denominator varies depending on the number of participants who completed the adverse events section of the questionnaire in each group.

^ Number (proportion) of participants reporting any related adverse event and number of each type of adverse event (number [proportion] of participants reporting each type of adverse event).

† Serious adverse events are defined as any untoward and unexpected medical occurrence that results in death, is life-threatening, requires hospitalisation or prolongation of existing inpatient's hospitalisation, results in persistent or significant disability or incapacity, is a congenital anomaly or birth defect, or any other important medical condition which may require medical or surgical intervention to prevent one of the outcomes listed.

‡ Defined as taken at least once per week over the prior month.

§ Defined as having tried the co-intervention specifically for their study hip pain in the previous 12 weeks at 3-months and in the previous 6 months at 9 months (but not including study interventions).

Appendix Table 1: Baseline characteristics of participants who provided both primary outcomes at 3-months and those participants who did not.

Characteristic	Incomplete primary outcomes (n=xxx)	Complete primary outcomes (n=xxx)
Group, n (%)		
1		
2		
Age, (years)		
Female, n (%)		
Height, (m)		
Body mass, (kg)		
Body mass index, (kg/m^2)		
Education level, n (%)		
Less than 3 years of high school		
3 or more years of high school		
Some education beyond high school		
Completed tertiary or higher education		
Current employment status, n (%)		
Currently employed		
Unable to work due to health reasons		
Retired (not due to health reasons)		
Unemployed/student		
Homemaker		
Hip symptom duration, (years)		
Unilateral symptoms, n (%)		
Pain type (PainDETECT) #		
Comorbid conditions, n (%)		
≥1 comorbid condition		
Heart disease		
High blood pressure		
Depression		
Anemia or other blood disease		
Ulcer or stomach disease		
Rheumatoid arthritis		
Diabetes		
Osteoarthritis		
Lung disease		
Cancer		

Back pain

Kidney disease

Liver disease

Other

Radiographic disease severity^, n (%)

Grade <2

Grade ≥2

Expectation of treatment outcome, n (%) ‡

No effect at all

Minimal improvement

Moderate improvement

Large improvement

Complete recovery

Treatments for hip in last 6 months, n (%)

≥1 treatment

Massage

Manual therapy

Gait aid

Ultrasound

Low level laser therapy (LLLT)

Transcutaneous Electrical Nerve Stimulation (TENS)

Hip braces

Heat/cold treatment

Land-based exercises (eg. strengthening, aerobic or stretching exercises)

Injections (eg. cortisone, Synvisc, platelet-rich plasma)

Hydrotherapy (eg warm water exercises)

Acupuncture

Arthroscopic surgery

Hip joint replacement surgery

Current pain medication use, n (%) *

≥1 medication used

Non-steroidal anti-inflammatory drugs

Acetaminophen

Topical anti-inflammatory drugs

Oral corticosteroids

Oral opioids

Pittsburgh Sleep Quality Index (PSQI)

Overall hip pain, past week (NRS)

Physical function, past 48 h (WOMAC)	
Hip pain during walking (NRS)	
Hip pain (WOMAC)	
Hip stiffness (WOMAC)	
Quality of life (AQoL-6D)	
Physical function (PSFS)	
Muscle strength (Nm/kg)	
30 s sit to stand (number)	
Timed stair climb (m/s)	
40 m fast walk (seconds)	
Submaximal cardiorespiratory fitness (ml/min)	

SD=standard deviation; kg=kilograms; m=metres; IQR= interquartile range (25th to 75th percentile); NRS=numerical rating scale, 0–10 with higher scores indicating more pain; WOMAC= Western Ontario and McMaster Universities Osteoarthritis Index, 0–68 for physical function, 0–20 for pain, 0–8 for stiffness with higher scores indicating more dysfunction/pain/stiffness; AQoL-6D=Assessment of Quality-of-Life instrument-6 dimension, -0.04–1.00 with higher scores indicating better quality of life; PSFS= Patient Specific Functional Scale, 0–10 with higher scores indicating better ability to perform self-selected activities at the same level as before hip problem.

The painDETECT questionnaire includes questions about pain course pattern, pain radiation and somatosensory phenomena. Total scores range from –1 to 38 with higher scores indicating more neuropathic-like symptoms.

^N = 127 participants (n = xx in Group 1; n = xx in Group 2). 69 missing due to COVID-19 lockdowns and restrictions in Australia from March 2020.

‡ Scored from a question asking about expectation of study treatment outcomes, with self-reported scores on a 5-point Likert scale from “0=no effect at all” to “4=complete recovery”.

* Defined as ≥ 1 time per week over the last 6 months.