Statistical Analysis Plan

A self-directed digital exercise program for hip osteoarthritis ("My Hip Exercise")

Statistical Analysis Plan

TRIAL FULL TITLE	A self-directed digital exercise program for hip osteoarthritis ("My Hip Exercise"): a randomised controlled trial.
TRIAL SHORT TITLE	MyHipExercise
TRIAL REGISTRATION	Australia New Zealand Clinical Trials Registry (ACTRN12622001533785).
PROTOCOL VERSION	v1 from 25/10/2022 named "My Hip Exercise protocol_v1_clean"
TRIAL CHIEF INVESTIGATORS	Prof Kim Bennell
TRIAL STATISTICIAN	Ms Fiona McManus, A/Prof Karen Lamb
SAP AUTHOR	Ms Fiona McManus, MISCH biostatistician, University of Melbourne A/Prof Karen Lamb (co-Head MISCH Biostatistics), University of Melbourne Prof Kim Bennell, Trial Chief Investigator, University of Melbourne
SAP VERSION	1
SAP VERSION DATE	21-Aug-25

SAP Revision History

Version	Reason(s) for change	Date
1.0	Initial version	21-Aug-25

SAP Signatures

I give my approval for the attached SAP entitled MyHipExercise_MISCH_SAPLong_v1 dated 21-Aug-25.

Chief Investigator

Name / Affiliation	Signature	Date
Prof Kim Bennell, University of Melbourne	[REDACTED]	25/8/2025

SAP Author

Name / Affiliation	Signature	Date
Ms Fiona McManus, University of Melbourne	[REDACTED]	25/8/2025

Senior Statistician

Name / Affiliation	Signature	Date
A/Prof Karen Lamb, University of Melbourne	[REDACTED]	21/8/25

Table of Contents

1 Introduction 7

1.1 Preface 7
1.2 Purpose of the analyses 7
2 Study Objectives and Endpoints 7
2.1 Study Objectives 7
2.2 Endpoints 7
3 Study Methods 8
3.1 General Study Design and Plan 8
3.2 Inclusion-Exclusion Criteria and General Study Population 10
3.3 Randomisation and Blinding 10
3.4 Study Variables 11
4 Sample Size 21
5 General Considerations 21
5.1 Timing of Analyses 21
5.2 Analysis Populations 22
5.2.1 Full Analysis Population 22
5.3 Covariates and Subgroups 22
5.4 Missing Data 22
5.5 Interim Analyses and Data Monitoring 23
5.6 Multiple Testing 23
5.7 Intercurrent Events 24
6 Summary of Study Data 24
6.1 Participant Disposition 24
6.2 Derived variables 27
6.3 Protocol Deviations 27
6.4 Demographic and Baseline Variables 27
6.5 Concurrent Illnesses and Medical Conditions 27
6.6 Prior and Concurrent Medications 27
6.7 Treatment Compliance 27
7 Efficacy Analyses 28
7.1 Primary Efficacy Analysis 28
7.2 Secondary Efficacy Analyses 29
7.3 Exploratory Efficacy Analyses 29
8 Safety Analyses 29
9 Reporting Conventions 30

10 Technical Details 30
11 Summary of Changes to the Protocol 31
12 References 31
13 Listing of Tables, and Figures 33
Appendix A – Planned tables and figure/s 34

Abbreviations and Definitions

ABBREVIATION	DEFINITION
AQoL-8D	Assessment of Quality of Life Instrument - 8 Dimension
ASES	Arthritis Self Efficacy Scale
ASGC	Australian Standard Geographical Classification
BFMS	Brief Fear of Movement Scale for Osteoarthritis
BMI	Body Mass Index
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
COX-2	Cyclooxygenase-2
DSMB	Data Safety Monitoring Board
EARS	Exercise Adherence Rating Scale
HOOS	Hip dysfunction and Osteoarthritis Outcome Score
ICH	International Council for Harmonisation
IPEQ-W	Incidental and Planned Exercise Questionnaire, version W
IQR	Interquartile Range
MCAR	Missing Completely At Random
MCID	Minimal Clinically Important Difference
MAR	Missing At Random
MNAR	Missing Not At Random
NICE	National Institute for Health and Care Excellence
NRS	Numerical Rating Scale
OA	Osteoarthritis
REDCap™	Research Electronic Data Capture
STATA-8D	Statistical Software for Data Analysis
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

1 Introduction

1.1 Preface

Hip osteoarthritis (OA) is a leading global cause of chronic pain and disability. Given there is no cure for OA, patient self-management is vital with education and exercise being core recommended treatments. However, there is under-utilisation of these treatments due to a range of clinician and patient factors. Innovative service models that increase patient accessibility to such treatments and provide support to engage are needed.

1.2 Purpose of the analyses

This study primarily aims to determine the effects of a self-directed digital exercise intervention comprising online education and exercise supported by a mobile app to facilitate adherence on the primary outcomes of changes in hip pain during walking and patient-reported physical function at 24-weeks when compared to online education control for people with hip OA.

2 Study Objectives and Endpoints

2.1 Study Objectives

To determine the effectiveness of a self-directed digital exercise program ("My Hip Exercise") comprising internet-delivered education, physical activity guidance and lower-limb strengthening exercise combined with a mobile app to monitor and support exercise participation on the primary outcomes of changes in hip pain during walking and patient-reported physical function at 24-weeks when compared to online education control for people with hip OA.

The primary hypothesis is that participants allocated to the group receiving the digital exercise program will have greater improvements in hip pain during walking and/or self-reported physical function at 24-weeks compared to those allocated to the control group.

Secondary objectives are to:

i) determine if the digital exercise program is more effective than control on other clinical outcomes (overall hip pain severity; function in sport and recreation; hip-related quality-of-life; health-related quality of life; global rating of overall change in the hip; arthritis self-efficacy for pain; exercise self-efficacy; physical activity levels; global change in physical activity; fear of movement; and use of oral pain medication) at 24 weeks; and

ii) determine engagement with, and usefulness of, the digital exercise program using a range of process measures.

iii) document the frequency and type of adverse events in each treatment arm.

2.2 Endpoints

The two primary endpoints are:

1. Change (baseline minus 24-weeks follow-up) in severity of hip pain while walking
2. Change (baseline minus 24-weeks follow-up) in physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

Secondary endpoints are:

1. Change (baseline minus 24-weeks follow-up) in pain subscale of Hip dysfunction and Osteoarthritis Outcome Score (HOOS)
2. Change (baseline minus 24-weeks follow-up) in function, sports and recreational activities subscale of HOOS
3. Change (baseline minus 24-weeks follow-up) in hip-related quality-of-life subscale of HOOS
4. Change (baseline minus 24-weeks follow-up) in health-related quality of life
5. Change (baseline minus 24-weeks follow-up) in physical activity level
6. Change (baseline minus 24-weeks follow-up) in fear of movement
7. Change (baseline minus 24-weeks follow-up) in pain self-efficacy
8. Change (baseline minus 24-weeks follow-up) in self-efficacy for exercise
9. Global rating of overall change in hip condition at 24-weeks
10. Global rating of change in physical activity at 24-weeks
11. Use of any oral pain medications for hip pain at least once a week in the prior month at 24 weeks
12. Achieved the minimal clinically important difference (MCID) in pain
13. Achieved the MCID in function

3 Study Methods

3.1 General Study Design and Plan

This study is a pragmatic two-arm parallel superiority double-blind randomised controlled trial (RCT) with an active control. Participants will be randomised with no stratification into one of two groups i) My Hip Exercise; or ii) Online education control.

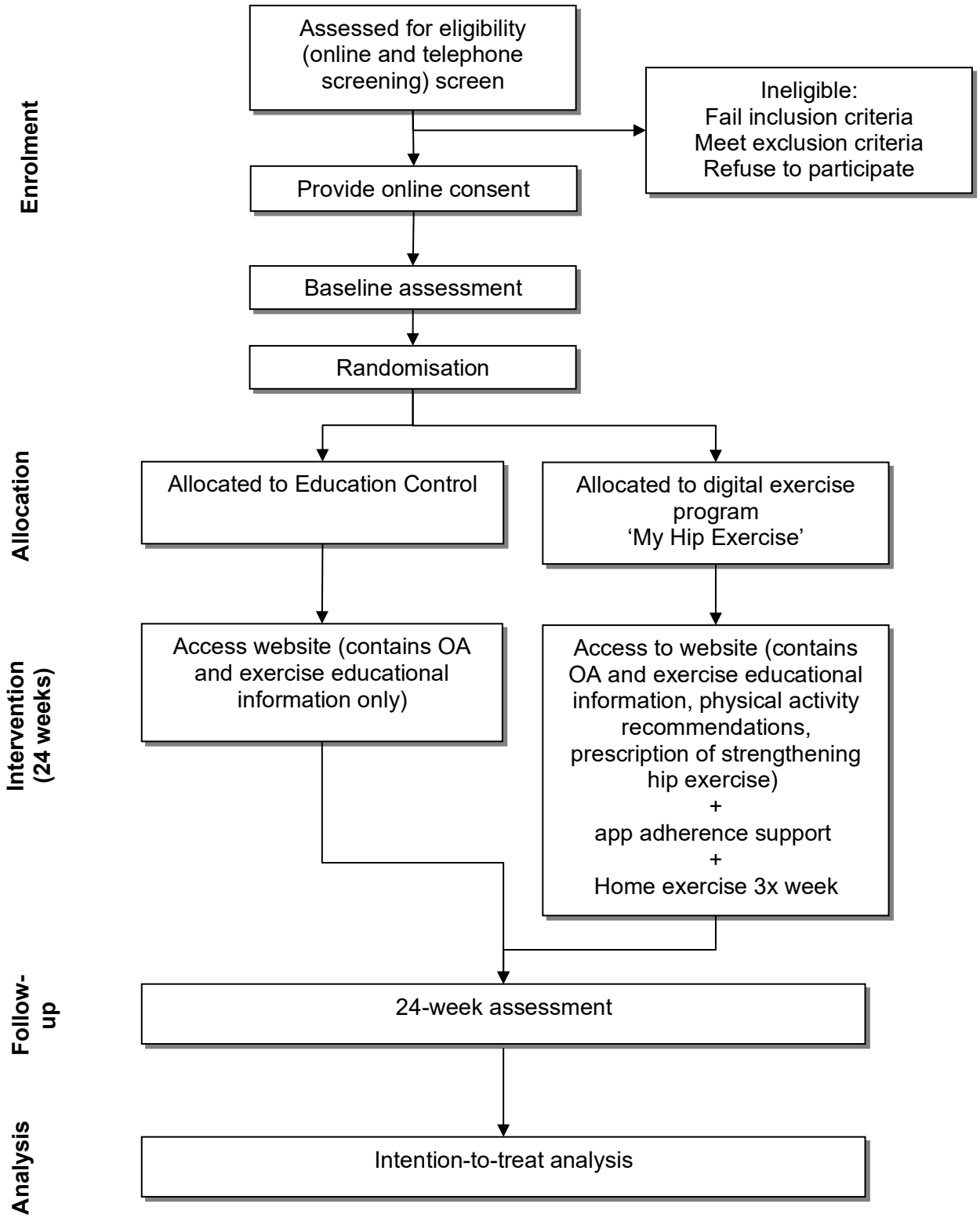


Figure 1. Study visits and procedures schedule.

3.2 Inclusion-Exclusion Criteria and General Study Population

Inclusion Criteria

- Meet the National Institute for Health and Care Excellence (NICE) osteoarthritis clinical criteria (1):
  - aged 45 years and over;
  - activity-related hip joint pain; and
  - hip joint morning stiffness ≤30 minutes or no hip joint morning stiffness
- Report hip pain for ≥ 3 months;
- Report hip pain on most days in the past month;
- Report average hip pain during walking in past week as ≥ 4 out of 10 on an 11-point numeric rating scale (NRS);
- Have a home Internet connection, a computer/tablet device that enables access to the Internet and a suitable phone to download an app; and
- Able to give informed consent and to participate fully in the interventions and assessment procedures.

Exclusion Criteria

- Have had a hip joint replacement in the most painful hip;
- Planning to undergo a hip joint replacement in the next 6 months;
- Have participated in regular leg strengthening exercise over the past 6 weeks (one or more times per week for each week);
- Self-reported diagnosis of rheumatoid arthritis or other inflammatory arthritis;
- Have had a fall within the last 12 months and do not receive clearance from a general practitioner to participate in an unsupervised home exercise program;
- Are housebound due to mobility limitations and requiring assistance from another person to leave the house in the previous month and do not receive medical clearance from a general practitioner to participate in an unsupervised home exercise program;
- Have a health condition(s) listed on the Exercise and Sports Science Australia stage 1 pre-exercise screening questionnaire that might compromise exercise safety (2) and do not receive medical clearance from a general practitioner to participate in an unsupervised home exercise program; and/or
- Unable to speak or read English.

3.3 Randomisation and Blinding

Participants will be randomised with no stratification into one of two groups: i) My Hip Exercise; or ii) Online education control. Computer-generated randomisation (1:1 ratio) will be prepared by an independent biostatistician in permuted random blocks of varying sizes. The schedule will be stored on a password-protected website (REDCap™) at the University of Melbourne and maintained by a researcher not involved in either participant recruitment or administration of outcome measures. Group allocation will be revealed by this same researcher after baseline assessment.

Due to the nature of the interventions, it is not possible to blind participants. As outcomes are self-reported measures, assessors are not blinded. However, limited disclosure will be used to ensure participants are blinded. Participants will be blinded to the alternative treatment and are informed at the time of recruitment that the study will investigate and compare digital resources/websites

aimed at helping people self-manage their hip pain and that this may include exercise and online resources. The statistical analysis plan will be written and published on our centre's website while blind to group allocation.

3.4 Study Variables

Measure name	Description	Scale	Time-points measured
Primary Outcome Measures*			
Severity of hip pain while walking	Scored on an 11-point NRS for "average amount of pain felt over the last week in your study hip when you are walking" where 0=no pain and 10=worst pain possible: nrs_walking_[0w=Baseline;24w=24 weeks].	Ranges from 0 to 10; higher scores indicate worse pain.	Baseline and 24-weeks
Physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (3)	Scored using 17 questions regarding hip function, with Likert response options ranging from no dysfunction=0 to extreme dysfunction=4. Sum 17 questions: hoosa1_[0w=Baseline;24w=24 weeks] to hoosa17_[0w=Baseline;24w=24 weeks].	Ranges from 0 (no dysfunction) to 68 (maximum dysfunction).	Baseline and 24-weeks
Secondary Outcome Measures*			
Pain subscale of Hip dysfunction and Osteoarthritis Outcome Score (HOOS) (4)	Scored using 10 questions regarding hip pain experienced in the last week, with 5-point Likert response options ranging from "None/Never=0" to "Extreme/Always=4": hoosp1_[0w=Baseline;24w=24 weeks] to hoosp10_[0w=Baseline;24w=24 weeks] (except for hoos_p1_24w].	Normalised to 0 – 100 scale (=100-(average of 10 questions)/4*100); higher scores indicate less pain.	Baseline and 24-weeks
Function, sports and recreational activities subscale of HOOS (4)	Scored using 4 questions about difficulty with sport and recreational activities in the last week, with 5-point Likert response options ranging from "none=0" to "extreme=4": hoossport1_[0w=Baseline;24w=24 weeks] to hoossport4_[0w=Baseline;24w=24 weeks].	Normalised to 0 – 100 scale (=100-(average of 4 questions)/4*100); higher scores indicating less difficulty.	Baseline and 24-weeks
Hip-related quality-of-life subscale of HOOS (4)	Scored using 4 questions about hip-related quality of life experienced in the last week, with 5-point Likert response options for each question ranging from "never/not at all/none =0" to "constantly/ totally/ extremely/ extreme =4":	Normalised to 0 – 100 scale (=100-(average of 4 questions)/4*100); higher scores indicate better hip-related quality-of-life.	Baseline and 24-weeks

	hoosqol1_[0w=Baseline;24w=24 weeks] to hoosqol4_[0w=Baseline;24w=24 weeks].		
Global rating of overall change in hip condition	Scored using a 7-point global rating of change Likert scale with response options ranging from “much worse=1” to “much better=7” when compared to baseline: grc_hip_[24w=24 weeks].	Participants indicating they are “moderately better=6” or “much better=7” will be classified as improved and all others as not improved. Number and proportion of participants improved or not improved will be reported.	24-weeks
Health-related quality of life	Assessment of Quality of Life Instrument (AQoL-8D)(5). Scored from 35 questions regarding health-related quality of life in the last week; calculated using https://www.monash.edu/business/che/aqol/using-aqol/scoring AQoL-8D utility algorithms, STATA-8D, accessed 7th May, 2024 and 1st April, 2025 (https://www.aqol.com.au/choice-of-aqol-instrument/58.html): aqol_q1_[0w=Baseline;24w=24 weeks] to aqol_q35_[0w=Baseline;24w=24 weeks] (except for aqol_q_19_[0w=Baseline;24w=24 weeks], aqol_q_24_[0w=Baseline;24w=24 weeks], aqol_q_26_[0w=Baseline;24w=24 weeks].	Ranges from -0.04 to 1.00; higher scores indicate better quality of life.	Baseline and 24-weeks
Physical activity level subscore – IPEQ planned sport activities	Incidental and Planned Exercise Questionnaire, version W (IPEQ-W)(6) total score unable to be derive due to errors in data collection of Q6 and Q8 therefore subscore, planned sport activities derived. Subscore reported on a scale of 0–42; higher scores indicate higher levels of activity: ex_class_[0w=Baseline;24w=24 weeks] (Q1) ex_class_length_[0w=Baseline;24w=24 weeks] (Q2) home_ex_[0w=Baseline;24w=24 weeks] (Q3) home_ex_length_[0w=Baseline;24w=24 weeks] (Q4) other_exyn_[0w=Baseline;24w=24 weeks] other_ex_number_[0w=Baseline;24w=24 weeks] ipeq_other1_[0w=Baseline;24w=24 weeks] other_ex1_number_[0w=Baseline;24w=24 weeks] other2_length_[0w=Baseline;24w=24 weeks] ipeq_other2_[0w=Baseline;24w=24 weeks] other_ex2_number_[0w=Baseline;24w=24 weeks] other2_length2_[0w=Baseline;24w=24 weeks]	IPEQ Planned sport activities uses Q1, Q2, Q3 and Q4 components and expressed as hours per week. The score is derived from multiplying the frequency score and duration score to create a Duration for the week score: planned sport activity = (Q1*Q2) + (Q3*Q4).	Baseline and 24-weeks

	ipeq_other3_[0w=Baseline;24w=24 weeks] other_ex3_number_[0w=Baseline;24w=24 weeks] other2_length3_[0w=Baseline;24w=24 weeks] ipeq_other4_[0w=Baseline;24w=24 weeks] other_ex4_number_[0w=Baseline;24w=24 weeks] other2_length4_[0w=Baseline;24w=24 weeks] ipeq2_[0w=Baseline;24w=24 weeks] (Q5) ipeq3_[0w=Baseline;24w=24 weeks] (Q6) ipeq4_[0w=Baseline;24w=24 weeks] (Q7) ipeq5_[0w=Baseline;24w=24 weeks] (Q8) ipeq6_[0w=Baseline;24w=24 weeks] (Q9) ipeq7_[0w=Baseline;24w=24 weeks] (Q10)		
Global rating of change in physical activity	Scored using a 7-point global rating of change Likert scale with response options ranging from “much less=0” to “much more=7” when compared to baseline: grcphysactivity_[24w=24 weeks].	Participants indicating that they are doing “moderately more=6” or “much more=7” will be classified as increasing physical activity. Number and proportion of participants increasing or not increasing will be reported.	24-weeks
Fear of movement	Brief Fear of Movement Scale for Osteoarthritis(7). Scored from 6 statements regarding fear of injury/re-injury due to movement on a 4-point Likert scale from 1 = “strongly disagree” to 4 = “strongly agree”; sum 6 scores: bfoms1_[0w=Baseline;24w=24 weeks] to bfoms6_[0w=Baseline;24w=24 weeks].	Scores range from 6 (minimal fear) to 24 (maximal fear).	Baseline and 24-weeks
Pain self-efficacy	Scored from the pain subscale of the Arthritis Self Efficacy Scale(8): Scored on a 10-point Likert scale from “very uncertain=1” to “very certain=10” for the 5 questions: ases_p1_[0w=Baseline;24w=24 weeks] to ases_p5_[0w=Baseline;24w=24 weeks].	Score is the mean of items in the scale from 1-10 with higher scores indicating greater self-efficacy.	Baseline and 24-weeks
Self-efficacy for exercise	Scored using the Self-Efficacy for Exercise Scale (9) which has 9 items scored on an 11-point NRS from “not confident=0” to “very confident=10”: see1_[0w=Baseline;24w=24 weeks] to see9_[0w=Baseline;24w=24 weeks].	Score is calculated by summing the responses to each question with a range from 0-90. A higher score indicates higher self-efficacy for exercise.	Baseline and 24-weeks
Use of oral pain medications for hip pain	Participants will self-report the use of any oral pain medication (yes/no), defined as one or more of oral non-steroidal anti-inflammatory drugs and/or analgesics (paracetamol combinations) and/or oral glucocorticoids (labelled Oral Corticosteroids)	Number and proportion of participants using any oral pain medication for hip pain at least once a	24-weeks

	and/or oral opioids taken at least once a week in the prior month for hip pain: anti_inflam_tablets_[24w=24 weeks] analgesia_paracet_[24w=24 weeks] oral_corticosteroids_ [24w=24 weeks] oral_opioids_[24w=24 weeks].	week in the prior month will be reported.	
Achieved the minimal clinically important difference (MCID) in pain	To aid clinical interpretation, the primary outcome, hip pain severity while walking, will be dichotomized into those who do and do not achieve the MCID in improvement in hip pain (1.8 NRS units): NRS_improved_24 = 1 if change (baseline minus follow-up) in walking hip pain at 24 weeks >= 1.8, zero otherwise.	Number and proportion of participants achieving the MCID in improvement in hip pain (1.8 NRS units) will be reported.	24-weeks
Achieved the minimal clinically important difference (MCID) in function	To aid clinical interpretation, the primary outcome, physical function subscale of the WOMAC, will be dichotomized into those who do and do not achieve the MCID in improvement in function (6 WOMAC units). WOMAC_improved_24 = 1 if change (baseline minus follow-up) in function at 24 weeks >= 6, zero otherwise.	Number and proportion of participants achieving the MCID in improvement in function (6 WOMAC units) will be reported.	24-weeks
Baseline descriptive measures			
Age	Calculated from date of birth: date_of_birth_[0w=baseline].	Reported in years.	Baseline
Sex	Self-reported via "What was your sex recorded at birth?" Male=1 Female=2 Another term (please specify)=3 sex (coding as above) sex_other_specify	Number and proportion of participants in each category will be reported.	Baseline
Gender	Self-reported via: "How do you describe your gender? Man or male=1 Woman or female =2 Non-binary =3 Prefer not to answer =4 I use a different term (please specify) =5	Number and proportion of participants in each category will be reported.	Baseline

	gender(coding as above) gender_otherterm		
Height	Self-reported baseline_height	Reported in metres between 1.47m to 2.01m.	Baseline
Weight	Self-reported baseline_weight	Reported in kgs between 40kg to 180kg.	Baseline
Body mass index (BMI)	Calculated as self-reported weight divided by the square of self-reported height.	Reported in kg/m^2.	Baseline
Ethnicity	Self-reported using: Australian/New Zealander=1 Aboriginal and Torres Strait Islander=2 European=3 Asian=4 Other Oceania=5 North African & Middle Eastern=6 Sub-Saharan African=7 North American =8 South American=9 Other (please list)=10 Ethnicity (coding as above) ethnicity_other	Number and proportion of participants in each category will be reported.	Baseline
Geographical location	Determined based on residential postcode and classified according to the Australian Standard Geographical Classification (ASGC) Remoteness Structure. Major cities = RA1 Inner regional = RA2 Outer regional = RA3 Remote = RA4 Very remote = RA5 Variable name in REDCap™: postcode Derived variable for georgraphic location: ASGS RA Code	Number and proportion of participants living in major cities, inner regional, outer regional, remote and very remote locations will be reported.	Baseline

Education level	Self-reported education level using a categorical scale with response options: Did not complete primary school=0; Primary school=1; Secondary school=2; Trade or trade certificate=3; University or tertiary institute=4; Higher university degree=5; Don't know/unsure=6. education	Number and proportion of participants in each category will be reported.	Baseline
Current employment status	Self-reported current employment status in response to the question: are you currently in paid employment (casual, part time or full time): yes=1, no=2. employment_status	Number and proportion of participants currently employed will be reported.	Baseline
Comorbidities	Reported using the Self-Administered Comorbidity Questionnaire (10). For all variables: yes=1, no=2. A prefix of "tre_" indicates the question "Do you receive treatment for it?" and a prefix of "lim_" indicates the question "Does it limit your activities?" Heart=Heart disease tre_heart lim_heart hbp=High blood pressure tre_hbp lim_hbp lung=Lung diease tre_lung lim_lung diab= Diabetes tre_diab lim_diab ulcer= Ulcer or stomach disease tre_ulcer lim_ulcer kidney= Kidney disease tre_kidney lim_kidney liver=Liver disease tre_liver lim_liver anemia=Anaemia or other blood disease tre_anemia lim_anemia cancer=Cancer tre_cancer lim_cancer dep=Depression tre_dep	Number and proportion of participants in each comorbidity will be reported, except Osteoarthritis and Rheumatoid Arthritis.	Baseline

	lim_dep osteo=Osteoarthritis tre_osteo lim_osteo back=Back pain tre_back lim_back ra=Rheumatoid arthritis tre_ra lim_ra other_prob=Other medical problems other_prob_desctx other_prob_1 other_prob_1_tre other_prob_1_lim other_prob_ckbx other_prob_2 other_prob_2_tre other_prob_2_lim		
Symptoms in other joints	Participants will be asked to tick joints from a list that they currently experience symptoms in. other_joints (1=I do not have problems in any other joints; 2=I have problems in other joints) problems_other_joints (1=Hand; 2=Neck, 3=Back, 4=Right knee, 5=Left knee, 6=Right foot/ankle, 7=Left foot/ankle, 8=Right shoulder, 9=Left shoulder)	Number and proportion of participants for each site will be reported.	Baseline
Duration of symptoms	Participants self-reported duration since their hip symptoms began. symptom_duration_years (0 to 81 [more than 80] years) sx_duration_months (0 to 11 months)	Reported in years as the sum of symptom_duration_years and (sx_duration_months/12).	Baseline
Medication usage	Participants will self-report the use of pain-relieving medications taken/used at least once a week in the prior month for hip pain (yes=1, no=2): anti_inflam_tablets_0w=Oral non-steroidal anti-inflammatory drugs cox2_tablets_0w=Oral non-steroidal anti-inflammatory cox-2 inhibitors analgesia_paracet_0w =Analgesia/paracetamol combinations topical_anti_inflam_0w =Topical anti-inflammatory gels or creams oral_corticosteroids_0w =Oral corticosteroids oral_opioids_0w =Oral opioids.	Number and proportion of participants using each medication for at least once a week in the prior month will be reported.	Baseline
Other measures			

Cointerventions	Participants will complete a custom-developed table to indicate the frequency of use (over the past 6 months) of a range of other hip treatments (yes=1, no=2): Massage massage_[24w=24 weeks] Manual therapy manual_therapy_[24w=24 weeks] Walking stick/gait aid walking_stick_[24w=24 weeks] Transcutaneous electrical nerve stimulation tens_[24w=24 weeks] Injection injections_[24w=24 weeks] Hydrotherapy hydrotherapy_[24w=24 weeks] Acupuncture acupuncture_[24w=24 weeks] Arthoscopic surgery arthroscopic_surgery_[24w=24 weeks] Hip joint replacement surgery tkr_[24w=24 weeks]	Participants who have used a treatment at least once in the past 6 months will be reported as a user of that treatment.	24-weeks
Adverse events	Related adverse events will be reported by participants using survey questions recorded in both Excel (over the 24-weeks) and REDCap (at 24-weeks) and will be defined as "any problem experienced in the study hip or elsewhere in the body deemed to be a result of participating in the trial AND at least one of i) caused negative/adverse symptoms/effects for two days or more, and/or ii) resulted in the participant seeking treatment or taking medication". Excel (yes=1): Hippain Shincalfpain Paininotherareas Anklefootpain Hipstiffness Kneepain Hipswelling These types of related adverse events were derived by CHESM from survey questions over 24-weeks and REDCap: adverse_[24w=24 weeks] (yes=1, no=0) If_ answered_yes: problem_1_nature problem_1_what_tx problem_1_health_prof problem_1_health_prof_type to_add_a_2nd_problem_click problem_2_nature problem_2_what_tx problem_2_health_prof problem_2_health_prof_type to_add_a_3rd_problem_click problem_3_nature problem_3_what_tx problem_3_health_prof	Number and proportion of participants who: - withdraw from the study due to a related adverse event - experience a serious related adverse event and types - experience one or more non-serious related adverse events and types.	24-weeks

	problem_3_health_prof_type to_add_a_4th_problem_click problem_4_nature problem_4_what_tx problem_4_health_prof problem_4_health_prof_type		
Process measures			
Leg strengthening exercise in the past 24 weeks (Control group only)	Participants will be asked "Over the past 24 weeks, did you perform leg strengthening exercise one or more times a week for at least 8 weeks?" With response options Yes=1 and No=0. strength_yn_[24w=24 weeks]	Number and proportion of participants reporting 'yes' will be reported.	24-weeks
Website visits	Self-reported. Participants will be asked "How many times did you visit the website over the past 2 months?" with responses "never"=0, "1-5 times"=1, "6-10 times"=2 and ">10 times"=3. website_use_[8w=8weeks;16w=16 weeks;24w=24 weeks]	Number and proportion of participants reporting each response will be reported.	8-weeks, 16-weeks, and 24-weeks
Downloads of Exercise program (Intervention group only)	Self-reported (yes=1, no=2) pdf1_yn pdf2_yn pdf3_yn	Number and proportion of participants that download exercise program PDFs 1, 2 and 3 will be reported.	24-weeks
Number of days home strengthening exercise performed (Intervention group only)	Participants will be asked "In the past two weeks, on how many days did you perform the strengthening exercises from the "My Hip Exercise" program (0-14). ex_number_mhe_[8w=8weeks;16w=16 weeks;24w=24 weeks]	Responses will be summed over the three time points and reported as the average number of days per week. 'Acceptable' adherence will be defined as performing the home strengthening exercise on an average of 2 days or more per week.	8-weeks, 16-weeks, and 24-weeks
Exercise adherence rating	Using the Exercise Adherence Rating Scale (EARS) Section B (11) - Scored from 6 questions rated on a 5-point Likert scale (0 - completely agree to 4 - completely disagree).	Ranges from 0 and 24. A higher score indicates better adherence.	24-weeks

(Intervention group only)	ears1_[24w=24 weeks] to ears6_[24w=24 weeks]	Items 1, 4 and 6 are reverse scored.	
Downloads of the "My Exercise Messages" app (Intervention group only)	Self-reported. Participants will report (Yes=1/No=0) if they downloaded the app. app_yn_[24w=24 weeks]	Number and proportion of participants who downloaded the app will be reported.	24-weeks
Frequency of entering exercise into the "My Exercise Messages" app (Intervention group only)	Self-reported. Scored on a 5-point Likert scale for "How often did you enter your number of exercise days into the app as requested" (Never=1, rarely=2, sometimes=3, often=4, always=5) app_ex_freq_[24w=24 weeks]	Number and proportion of participants for each response.	24-weeks
Use of the "My Exercise Messages" app (Intervention group only)	Participants will be asked "On how many days in the past 14 days did you use the "My Exercise Messages" app (including reading any notifications, entering data and/or opening the app)" with response options: "Never"=0, "1-5 times"=1, "6-10 times"=2 and ">10 times"=3. app_use_[8w=8weeks;16w=16 weeks;24w=24 weeks]	Number and proportion of participants for each response.	8-weeks, 16-weeks, and 24-weeks
Usefulness of the "My Hip Exercise" website (Intervention group only)	Self-reported. Scored on an 11-point NRS for "How useful did you find the "My Hip Exercise" website program? (not including the app)" from 0 = "not at all useful" to 10 = "extremely useful". website_useful_[24w=24 weeks]	Ranges from 0 to 10; higher scores indicate greater perceived usefulness.	24-weeks
Usefulness of the "My Exercise Messages" app (Intervention group only)	Self-reported. Scored on an 11-point NRS for "How useful did you find the "My Exercise Messages" app?" from 0 = "not at all useful" to 10 = "extremely useful". app_useful_[24w=24 weeks]	Ranges from 0 to 10; higher scores indicate greater perceived usefulness.	24-weeks
Exercise equipment use	Self-reported. Did you use exercise equipment (such as weights or elastic bands) to add resistance to	Number and proportion of participants using	24-weeks

(Intervention group only)	your leg for your strengthening exercises? (Yes=1/No=0) equipment_yn _[24w=24 weeks]	exercise equipment will be reported.	
Registration to recommended external websites (Intervention group only)	Self-reported via 1. "The My Hip Exercise website included a link to an online course "Taking Control of Your Hip and Knee Osteoarthritis" for people who want to learn more about osteoarthritis. Did you register for this course? Yes=1/No=0 [include screenshot of course] oa_course_yn 2. "The My Hip Exercise website included a link to an online course "PainTRAINER" for people who want to learn strategies to manage pain. Did you register for this course? Yes=1/No=0 [include screenshot of course] paintrainer_yn	Number and proportion of participants reporting "Yes" will be reported.	24 weeks

*All continuous primary and secondary outcomes will be calculated as change in (baseline minus 24-week follow up).

4 Sample Size

The sample size is based on detecting a difference in change between-groups that meets or exceeds a specified minimal clinically important difference (MCID) for the two primary outcomes of hip pain on walking and WOMAC physical function. The MCIDs we are using are 1.8 units for NRS hip pain (12) and 6 units for WOMAC physical function (13). To achieve 80% power and a two-sided 5% significance level split equally across the two primary outcomes, assuming equal between-participant standard deviations of 2.5 for pain and 13 for function for both groups and a correlation between pre- and post-measurements of 0.25 for pain and 0.40 for function (14), and accounting for 15% loss to follow up, we require 91 participants per arm, for a total of 182 participants. We will consider the intervention to be effective if at least one of the two primary outcomes shows a significant between-group difference.

5 General Considerations

5.1 Timing of Analyses

The final analysis will be performed after all randomised participants have had the opportunity to complete final data collection on data transferred to the file "MHEx Age RA Code AE Data for Stats" and "MyHipExerciseQuestio_DATA_NOHDRS_2025-MM-DD_TTTT", having been documented as meeting the cleaning and approval requirements of "SOP03 CHESM Data Collection and Storage_v1.0" and "MyHipExercise_MISCH_Data Cleaning_V1" and after the finalisation, approval and publication on our research centre's website of this SAP document.

Study data entered and stored in the study database will be transferred to the trial biostatistician. After the last participant has concluded study participation, all study data are available, and all study data have been cleaned, a blinded review will be undertaken with the subgroup of the study team consisting of the principal investigator, project manager and study biostatisticians prior to database lock. During the blinded data review meeting, the following topics will be discussed and decided upon without knowledge of the underlying treatment code:

1. Participants who have withdrawn consent, in relation to the use of the participant's data (or part of it) in any of the analyses.
2. Participant's inclusion or exclusion status with regards to the "as randomised" and "as treated" treatment group (e.g., in the case of misallocation). The exact process for assigning the statuses will be defined and documented in the minutes of the meeting prior to unblinding to study name along with the reason for classifying "as randomised" and "as treated".
3. Participant's inclusion or exclusion status with regards to each analysis population.
4. Participant's intercurrent events, in relation to the intercurrent event handling in the analysis of the participant's data (or part of it). Intercurrent events will be listed and tabulated for the purpose of the blinded data review meeting.

After database lock, the randomisation code will be obtained from the independent statistician who generated the randomisation code and added to the study database by the trial statistician. No database may be locked, randomisation code unmasked, or analyses completed until the SAP has been approved (in case of updates since its initial approval). The planned analyses in this SAP will be conducted after unmasking of the randomisation code and any changes to this SAP after unmasking will be documented in an amendment of this SAP and considered post-hoc analyses.

5.2 Analysis Populations

5.2.1 Full Analysis Population

The full analysis population will be all participants who were randomised and who did not withdraw all their data, as documented in the minutes of the blinded data review meeting. This population will be used for all data including background, safety and efficacy data.

5.3 Covariates and Subgroups

For continuous outcomes, linear regression models will adjust for the baseline measurements of the relevant outcome.

No variables were used to stratify randomisation and so no stratifying variables will be adjusted for in any analyses.

No subgroup analyses will be conducted.

5.4 Missing Data

If missing data are present, an appendix table will provide summaries of baseline characteristics and baseline levels of primary and secondary outcomes where measured between two groups: those participants who provide both primary outcomes post-intervention, and those participants who are missing either or both primary outcomes.

If less than 5% of data both primary outcomes are missing, missing data will be assumed Missing Completely At Random (MCAR). This means the cause of the data being missing is unrelated to the missing or observed data and that analyses of complete data should provide an unbiased estimate of the effect that would be estimated should all participant data be available. Linear regression analyses of primary and secondary outcomes will be performed on complete case data (i.e., participants with missing data will be dropped from analyses).

Should the amount of missing data for either primary outcome be greater than 5%, missing data will be assumed Missing At Random (MAR). This means we will assume that differences between missing and observed values can be explained by other observed values, such as baseline variables. Multiple imputation will be conducted to impute data for all outcomes under this assumption. The primary analysis for all outcomes will then use multiply imputed datasets, with a sensitivity analysis using complete case data. Multiple imputation will still be valid if data are actually MCAR and not MAR. If data are actually MCAR, multiple imputation will improve precision by increasing the sample size.

Missing baseline characteristics will be imputed using single mean imputation for continuous variables and imputation of the most frequent value for categorical variables. Missing outcomes at follow-up will be imputed using chained equations with predictive mean matching and five nearest neighbours for continuous and binary outcomes. Imputation models for outcomes will include all primary and secondary outcomes at both baseline and post-intervention timepoints, along with age, sex, BMI, ethnicity, geographical location, education level, employment status, comorbidities, symptoms in other joints, duration of hip symptoms and medication use. Data will be imputed for each treatment group separately. The number of imputed data sets created will be based on the percentage of participants in the sample with missing outcome data (e.g., 15 imputed datasets if 15% of participants have missing data) or a minimum of 10 imputed datasets, whichever is greater. 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 outcomes. Estimates from the imputed datasets will be combined using Rubin's rules (15).

Results using complete case datasets (that assume MCAR) and multiply imputed datasets that assume missing data is MAR may be biased if missing data are actually Missing Not At Random (MNAR), which means that data that is missing is missing due to the missing data value itself. If more than 5% of either primary outcome is missing, then sensitivity analyses will be conducted using the delta-adjustment method under the pattern-mixture modelling framework in the context of multiple imputation to assess sensitivity to MNAR for the primary outcomes with a range of plausible delta parameters for the two primary continuous outcomes (16).

5.5 Interim Analyses and Data Monitoring

No interim analyses will be performed. As this was a low-risk trial, no data safety monitoring board (DSMB) was required. Adverse events were self-reported by participants at 24-weeks follow-up.

5.6 Multiple Testing

The two-sided 5% significance level will be split equally across the two primary outcomes (alpha level of 0.025 each primary outcome): adjusted 95% confidence intervals and adjusted p-values will be reported for the two primary outcomes.

We have several secondary outcomes, which are exploratory and not powered for. We will therefore not adjust for multiple secondary outcomes; instead, the estimates and unadjusted 95% CIs will be reported in order to let readers use their own judgment about the relative weight of the conclusions regarding the effect of the "My Hip Exercise" digital exercise program on secondary outcomes. This approach aligns with the approach favoured by the American Statistical Association (17).

5.7 Intercurrent Events

Intercurrent events are events that occur post-randomisation and may preclude the observation of the outcome variable or affect its measurement. Intercurrent events will be tracked by the study team during the study. Data will be reviewed prior to the closure of the database to ensure all intercurrent events are captured.

Possible intercurrent events include the following:

- Adverse events
- Pain medication
- Cointerventions
- Hip surgery

Intercurrent events will be summarised by each intervention group.

6 Summary of Study Data

All approximately symmetrically distributed continuous variables will be summarised using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, while other continuous variables will be summarised as n (non-missing sample size), median, 25th and 75th percentiles (interquartile range). The count and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by treatment. All summary tables will be structured with a column for each treatment in the order (Control, Intervention) and will be annotated with the total population size relevant to that table/treatment, including any missing observations, if relevant, in either the table or a footnote of the table.

The sample size of non-missing values for univariable summary statistics may be larger than the sample size of non-missing values in a complete-case analysis used in the primary regression analysis, as different participants may have missing values in different variables, such as baseline covariates and the endpoints.

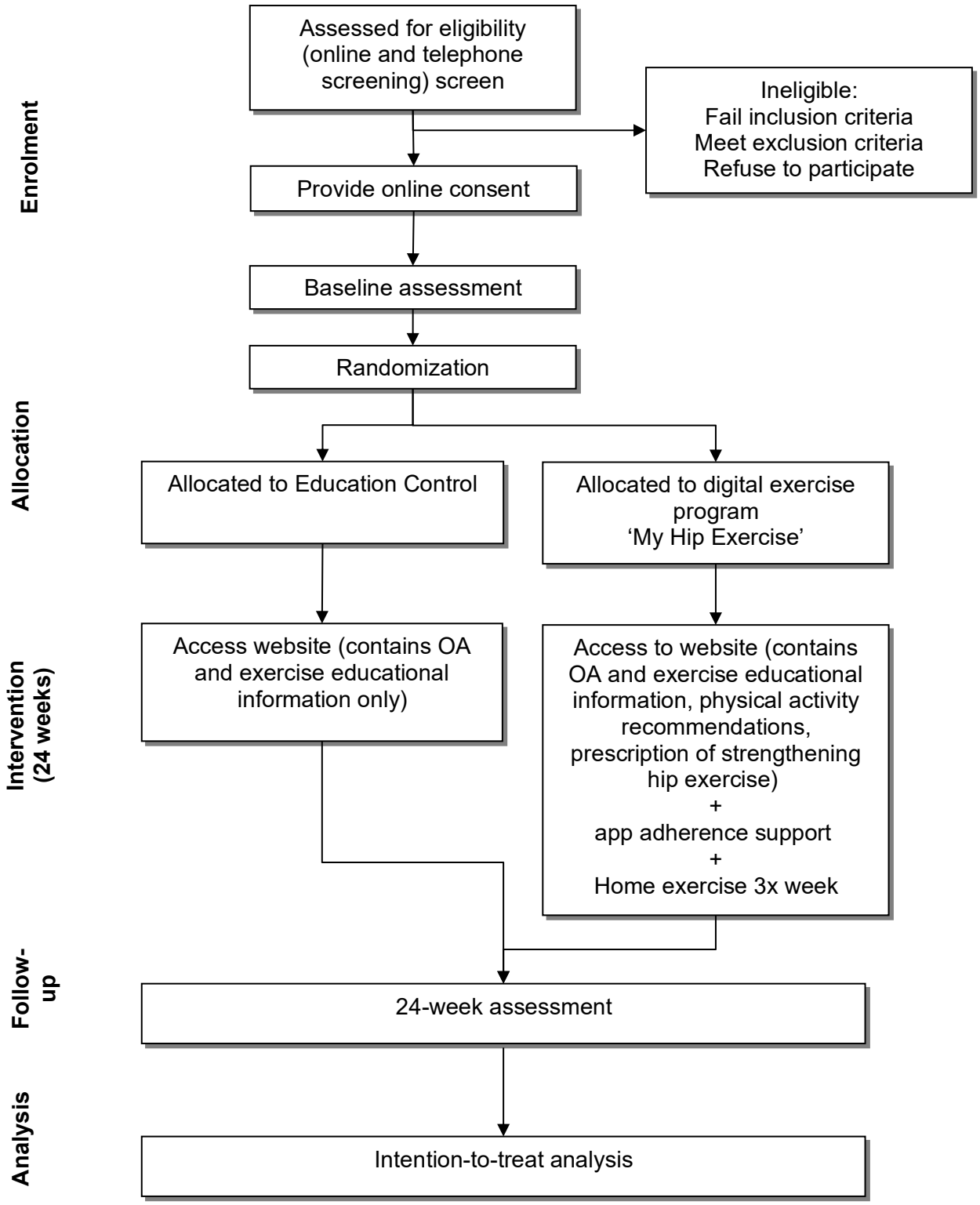
6.1 Participant Disposition

The flow of participants will be presented in a Consolidated Standards of Reporting Trials (CONSORT) diagram, according to the CONSORT statement and relevant extensions (18). The following will be reported:

- Number of participants assessed for eligibility (count of relevant variable from MHEx Recruitment Database)
- Number of participants ineligible: sum of number failed inclusion criteria (count of relevant variable from MHEx Recruitment Database), number met exclusion criteria count of relevant

variable from MHEx Recruitment Database), and declined participation (count of '2'=no for 'confirmation' variable from REDCap™ My Hip Participant Details form or count of relevant variable from MHEx Recruitment Database) and any other reasons (relevant variable from MHEx Recruitment Database)

- Number of participants providing online consent (count of '1'=yes for 'confirmation' variable from REDCap™ My Hip Participant Details form or count of relevant variable from MHEx Recruitment Database)
- Number of participants completing baseline assessment (count of '2'=complete for 'my_hip_baseline_questionnaire_complete' variable from REDCap™ My Hip Baseline Questionnaire form)
- Number of participants randomized (count of '1'=yes for 'randomised' variable in My Hip Participant Details form)
- Number of participants allocated to each of the two intervention arms (count for each group number in 'group_allocation' variable from REDCap™ My Hip Participant Details form)
- Number of participants at follow-up and number of participants who withdrew from the study ('withdrawn' variable from REDCap™ My Hip Participant Details form) or who were lost to follow-up (count of missing values of 'nrs_walking_24w' or 'hoosa124w' to 'hoosa17_24w' variables from REDCap™ My Hip 6-month Questionnaire form, whichever is less)
- Number of participants included in the intention-to-treat analysis of each primary outcome (count of non-missing values of 'nrs_walking_24w' and 'hoosa1_24w' to 'hoosa17_24w' variables from REDCap™ My Hip 6-month Questionnaire form if not missing outcome at baseline ('nrs_walking_0w' and 'hoosa1_0w' to 'hoosa17_0w' variables), nor 'group_allocation')


Enrolment
Assessed for eligibility (online and telephone screening) screen
Ineligible:
Fail inclusion criteria
Meet exclusion criteria
Refuse to participate
Provide online consent
Baseline assessment
Allocation
Randomization
Allocated to Education Control
Allocated to digital exercise program
'My Hip Exercise'
Intervention (24 weeks)
Access website (contains OA and exercise educational information only)
Access to website (contains OA and exercise educational information, physical activity recommendations, prescription of strengthening hip exercise)
+
app adherence support
+
Home exercise 3x week
Follow-up
24-week assessment
Analysis
Intention-to-treat analysis

6.2 Derived variables

All outcomes are derived variables (See Section 3.4).

6.3 Protocol Deviations

Any protocol deviations including errors applying inclusion/exclusion criteria and/or administration of the wrong intervention will be documented, if they occur, in the findings manuscript with the trial results (patient flow diagram/text). The number of participants who were allocated the incorrect intervention will be reported overall and by treatment arm. All participants will be analysed according to randomised treatment arm in the intention-to-treat analysis, irrespective of these deviations. In sensitivity analysis, participants will be analysed according to the treatment they received, irrespective of what they were randomised to. Sensitivity analysis will be performed only for the primary outcomes.

6.4 Demographic and Baseline Variables

Baseline variables are listed in section 3.4. There are no deviations in presentation of baseline variables from section 6. Demographic and baseline characteristics of participants will be inspected to assess baseline comparability of treatment groups. No p-values will be reported to compare baseline variables between the two treatment arms.

6.5 Concurrent Illnesses and Medical Conditions

The coding system for comorbidities is as per section 3.4. Participants will report at baseline any comorbidities and will report any adverse events at 24 weeks. The information will be collated in the descriptive statistics table for baseline data, and adverse events table for adverse event data at 24 weeks. No p-values will be reported to compare variables between the two treatment arms.

6.6 Prior and Concurrent Medications

The coding system for medications and cointerventions is as per section 3.4. Participants will report at baseline and follow-up (24 weeks) if they were receiving any oral pain medications and any co-interventions at 24 weeks follow-up. This information will be collated in the descriptive statistics table, the binary outcomes table and the medications/cointerventions table. 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 listed in section 3.4 once in the past 6 months will be reported as a user. This will be summarised by control / intervention group. If there is missing data, the non-missing sample will be reported either in the table or in the footnote of the table. No p-values will be reported to compare the two treatment arms.

6.7 Treatment Compliance

Process variables, including the Exercise Adherence Rating Scale and the number of days (times) home strengthening exercises were performed, are listed and detailed in section 3.4. Data from these process measures will be reported using descriptive statistics as appropriate for each treatment group in the adherence table. In this study, participants in the My Hip Exercise digital exercise program group were advised to perform home strengthening exercises 3 times/week (days/week). 'Acceptable' adherence will be defined as performing the home strengthening exercise on an average of 2 days or more per week. The average number of days that the home strengthening exercises are performed will be calculated by averaging responses over the three time points (8 weeks, 16 weeks and 24 weeks) and reported as the average number of days per week.

Where data is missing from one or two timepoints, the average will be calculated from the remaining timepoints. Where data is missing for all 3 timepoints, non-compliance will be assumed.

7 Efficacy Analyses

A Statistical Analysis Plan will be written and published on our research centre's website prior to commencement of data analysis while blind to group allocation. Analyses comparing the two groups will be performed by the statistician blind to group details according to the intention-to-treat principle.

7.1 Primary Efficacy Analysis

The primary estimand is defined according to the addendum to the ICH E9 on estimands and sensitivity analysis in clinical trials (19). My Hip Exercise aims to answer the specific research questions: does a self-directed digital exercise program ("My Hip Exercise") comprising internet-delivered education, physical activity guidance and lower-limb strengthening exercise combined with a mobile app to monitor and support exercise participation improve the primary outcomes of i) change in hip pain severity during walking and/or ii) change in WOMAC physical function at 24-weeks when compared to online education control for people with hip OA?

Primary estimand attributes:

- Treatment: "My Hip Exercise" digital exercise program (intervention) compared to online education (control), allocation by randomisation.
- Population: Adults who meet the National Institute for Health and Care Excellence clinical criteria for OA, be aged 45 years and over, have activity-related hip joint pain, hip joint morning stiffness ≤30 minutes or no hip joint morning stiffness, report hip pain for ≥ 3 months, report hip pain on most days in the past month, report average hip pain during walking in past week as ≥ 4 out of 10 on an 11-point numeric rating scale (NRS), have a home Internet connection, a computer/tablet device that enables access to the Internet and a suitable phone to download an app; and be able to give informed consent and to participate fully in the interventions and assessment procedures.
- Variables: i) hip pain severity during walking at 24 weeks post-randomisation; ii) WOMAC physical function at 24 weeks post-randomisation.
- Intercurrent events: Possible intercurrent events outlined in Section 5.6 will be handled using the treatment policy strategy.
- Population level summaries: i) mean difference in change (baseline minus follow up) in hip pain severity during walking between arms ("My Hip Exercise" digital exercise program vs. online education control) and ii) mean difference in change (baseline minus follow up) in WOMAC physical function between arms ("My Hip Exercise" digital exercise program vs. online education control).

The number and percentage of participants who had intercurrent events will be described overall and by treatment arm in either text or relevant tables.

The primary hypothesis is that participants allocated to the group receiving the digital exercise program will have greater improvements in hip pain during walking and/or self-reported physical function at 24-weeks compared to those allocated to the control group.

For the continuous primary outcomes, mean differences in change (baseline minus follow up) will be compared between groups using separate linear regression models for each outcome adjusted for baseline levels of the relevant outcome. Model assumptions will be assessed using standard diagnostic plots.

The primary hypotheses will be evaluated using the estimate, multiplicity adjusted two-sided 95% confidence interval, and the multiplicity adjusted p-value of the i) mean difference in change (baseline minus 24-week follow up) in severity of hip pain while walking between the two treatment arms and ii) mean difference in change (baseline minus 24-week follow up) in WOMAC physical function between the two treatment arms.

If multiple imputation is used in the primary analysis, sensitivity analysis for the primary outcomes will be undertaken using available case analysis.

7.2 Secondary Efficacy Analyses

The following secondary outcomes will be treated as continuous outcomes and will be analysed similarly to the primary outcomes, and targeting a similar primary estimand: changes from baseline to 24-weeks follow-up in the HOOS pain subscale, HOOS function, sports and recreational activities subscale, HOOS hip-related quality-of-life subscale, health-related quality of life, physical activity level, fear of movement, pain self-efficacy, and self-efficacy for exercise.

For continuous secondary outcomes, mean differences in change (baseline minus follow up) will be compared between groups using separate linear regression models for each outcome adjusted for baseline levels of these outcomes. Model assumptions will be assessed using standard diagnostic plots.

For binary outcomes (global ratings of change in hip condition and physical activity, use of oral pain medications, and achievement of minimal clinically important differences in pain and function), groups will be compared using risk differences and risk ratios, calculated from logistic regression models.

Unadjusted 95% confidence intervals will be reported for all secondary outcomes.

7.3 Exploratory Efficacy Analyses

It is anticipated that not all participants will adhere to the intervention. Therefore, the analysis undertaken for the primary estimand will only provide an estimate of the causal effect of intervention assignment rather than an estimate of the causal effect of the intervention actually received. A sensitivity analysis will estimate treatment effects on the primary outcomes at 24-weeks assuming ‘acceptable’ adherence to the home strengthening exercise program (‘acceptable’ adherence defined as performing home strengthening exercises on an average of 2 or more days per week). Complier average causal effects will be estimated using an instrumental variables approach (where randomization is the instrument for adherence). Two-stage least squares models will be fitted (20) with complier average causal effects reported with unadjusted 95% confidence intervals.

8 Safety Analyses

In this trial, related adverse events are defined as "any problem experienced in the study hip or elsewhere in the body deemed to be a result of participating in the trial and at least one of i) caused negative/adverse symptoms/effects for two days or more, and/or ii) resulted in the participant

seeking treatment or taking medication". Adverse events will be ascertained by survey questions to participants at 24 weeks. The following will be reported:

- the number (and proportion) of participants reporting one or more non-serious related adverse events for each treatment group as outlined in the full analysis population section and the number (and proportion) of participants reporting each type of non-serious related adverse event;

- the number and proportion of participants that withdraw from the trial due to a related adverse event (if a participant(s) has/have notified us of this);

- the number (and proportion) of participants that experience one or more serious related adverse events and the number (and proportion) of participants reporting each type of serious related adverse event (defined as any untoward medical occurrence that i) results in death; ii) is life-threatening; iii) requires hospitalisation or prolongation of existing inpatient hospitalisation; iv) results in persistent or significant disability or incapacity; v) is a congenital anomaly or birth defect, or; vi) any other important medical condition which, although not included in the above, may require medical or surgical intervention to prevent one of the outcomes listed).

9 Reporting Conventions

The standard deviation will be reported with mean values, and interquartile range (25th and 75th percentiles) with the median. Percentages will be reported without decimal points, and percentages less than 1% will be presented as <1%.

P-values ≥0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, interquartile range or minimum and maximum will use the same number of decimal places as the original data. In any cells in a row of data for which no data are reported, NA (Not Applicable) or two mid-dots (··) will be included in the table and NA (Not Applicable) as a footnote.

For publication purposes, if the target journals guidelines differ in their reporting conventions to those listed above, the reporting will change to reflect the guidelines of the journal.

10 Technical Details

Analysis will be conducted using Stata/SE for Windows version 18.

A second review statistician will independently reproduce the primary analyses and the table titled "Change from baseline within groups on continuous outcomes across time from baseline, and difference in change between groups". The reviewing statistician will have an overview of the entire analyses and will explicitly check the code producing the other tables (selected at random) as well as any other pieces of code as desired.

To provide high quality code that is understandable, and allows reproduction of the analysis, a log capturing the version of the software and any external add-on code will be used.

Also, at the start of any code file there will be a set of comments that give

- the author
- the date of writing
- references to inputs and outputs

11 Summary of Changes to the Protocol

The following changes to the protocol are made in this statistical analysis plan:

- Intercurrent Events: The reporting of intercurrent events is detailed in the SAP.
- Estimand Framework: The use of the estimand framework to define how the treatment effect will be estimated is outlined in the SAP.
- Incidental and Planned Exercise Questionnaire, version W (IPEQ-W)[6] total score at 24 weeks is unable to be derived due to errors in data collection therefore the subscore, planned sport activities, will be derived and will be presented as a substitute for the IPEQ-W total score secondary outcome measure.

12 References

1. National Clinical Guideline Centre. Osteoarthritis: care and management in adults. 2014.
2. Exercise and Sports Science Australia. Adult Pre-Exercise Screening System (APSS). 2019.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. Journal of Rheumatology. 1988.
4. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. BMC musculoskeletal disorders. 2003;4(1):10.
5. Osborne RH, Hawthorne G, Lew EA, Gray LC. Quality of life assessment in the community-dwelling elderly: validation of the Assessment of Quality of Life (AQoL) Instrument and comparison with the SF-36. Journal of clinical epidemiology. 2003;56(2):138–47.
6. Delbaere K, Hauer K, Lord SR. Evaluation of the incidental and planned activity questionnaire for older people. British journal of sports medicine. 2010;44(14):1029–34.
7. Shelby RA, Somers TJ, Keefe FJ, DeVellis BM, Patterson C, Renner JB, et al. Brief fear of movement scale for osteoarthritis. Arthritis care & research. 2012;64(6):862–71.
8. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum. 1989;32(1):37–44.
9. Resnick B, Jenkins LS. Testing the reliability and validity of the Self-Efficacy for Exercise scale. Nursing research. 2000;49(3):154–9.
10. 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 Care & Research: Official Journal of the American College of Rheumatology. 2003;49(2):156–63.
11. Newman-Beinart NA, Norton S, Dowling D, Gavriloff D, Vari C, Weinman JA, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy. 2017;103(2):180–5.
12. Bellamy N, Carette S, Ford P, Kean W, Le Riche N, Lussier A, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials--results of a consensus development (Delphi) exercise. The Journal of rheumatology. 1992;19(3):451–7.
13. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Care & Research: Official Journal of the American College of Rheumatology. 2001;45(4):384–91.

14. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, 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):1833–42.
15. Carpenter JR, Bartlett JW, Morris TP, Wood AM, Quartagno M, Kenward MG. Multiple imputation and its application: John Wiley & Sons; 2023.
16. Rezvan PH, Lee KJ, Simpson JA. Sensitivity analysis within multiple imputation framework using delta-adjustment: Application to Longitudinal Study of Australian Children. Longitudinal and Life Course Studies. 2018;9(3):259–78.
17. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p< 0.05". Taylor & Francis; 2019. p. 1–19.
18. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, Group* CN. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Annals of internal medicine. 2017;167(1):40–7.
19. Committee for Medicinal Products for Human Use. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Amsterdam, the Netherlands: European Medicines Agency; 2020.
20. Stuart EA, Perry DF, Le H-N, Ialongo NS. Estimating intervention effects of prevention programs: Accounting for noncompliance. Prevention Science. 2008;9(4):288–98.

13 Listing of Tables, and Figures

Table 1. Baseline characteristics of participants, by intervention group, reported as mean (standard deviation) unless otherwise stated.

Table 2. Continuous outcome measures across time, by treatment group.

Table 3. Change in continuous outcomes within and between groups.

Table 4. Binary secondary outcomes.

Table 5. Process measures.

Table 6. Adverse events and co-interventions at 24 weeks.

Table 7. Effects of the intervention on the primary outcomes at 24-weeks assuming acceptable adherence (with acceptable adherence defined as performing the home strengthening exercise on an average of 2 days or more per week).

NOTE: The population set for all tables will be the 'as randomised' full analysis population.

Figure 1. Flow of participants through the trial.

Appendix A – Planned tables and figure/s

Table 1: Baseline characteristics of all randomized participants, by intervention group, reported as mean (standard deviation) unless otherwise stated.

	Control group (n=xxx)	Intervention group (n=xxx)
Age (years)		
Sex, n (%)		
Female		
Male		
Other		
Gender, n (%)		
Woman or female		
Man or male		
Non-binary		
Prefer not to answer		
I use a different term		
Weight (kg), median [IQR]		
Height (m)		
Body mass index (kg/m^2), median [IQR]		
Ethnicity, n (%)		
Australian/New Zealander		
Aboriginal and Torres Strait Islander		
European		
Asian		
Other Oceanian		
North African & Middle Eastern		
Sub-Saharan African		
North American		
South American		
Other		
Geographical location, n (%)[a]		
Major cities		
Inner regional		
Outer regional		
Remote		
Very remote		
Education level, n (%)		
Did not complete primary school		
Primary school		
Secondary/high school		
Trade or trade certificate		
University or tertiary institute		
Higher university degree		
Don't know/unsure		
Currently in paid employment, n (%)		
Problem in other joint, n (%)		
≥1 problem in other joint		
Hand		
Neck		
Back		
Knee		

Foot/ankle	
Shoulder	
Hip symptom duration (years), median [IQR]	
Comorbidities, n (%)[b]	
≥1 comorbid condition	
Heart disease	
High blood pressure	
Lung disease	
Diabetes	
Ulcer or stomach disease	
Kidney disease	
Liver disease	
Anaemia or other blood disease	
Cancer	
Depression	
Back pain	
Other	
Current pain medication use, n (%)[c]	
≥1 medication used	
Non-steroidal anti-inflammatories	
COX-2 inhibitors	
Analgesia (paracetamol combinations)	
Topical anti-inflammatories	
Oral corticosteroids	
Oral opioids	

SD=standard deviation; IQR=interquartile range; COX-2= cyclooxygenase-2.
[a] Based on residential postcode, in accordance with Australian Statistical Geography Standard.
[b] Comorbid conditions assessed using the Self-Administered Comorbidity Questionnaire.
[c] Defined as used at least once per week over the past four weeks for hip pain.

Table 2. Continuous outcome measures across time, by treatment group[a].

	Baseline		24 weeks	
	Control (n=xx)	**Intervention (n=xx)**	**Control (n=xx)[b]**	**Intervention (n=xx)[c]**
Primary outcomes				
Hip pain severity during walking (NRS)[d]				
Physical function (WOMAC)				
Secondary outcomes				
Hip dysfunction and Osteoarthritis Outcome Score[e]				
Pain				
Function in sport and recreation				
Hip-related quality of life				
Health-related quality of life (AQoL-8D)[f]				
Physical activity (IPEQ-W)				
Fear of movement (BFMS)				
Pain self-efficacy (ASES pain subscale)				
Self-efficacy for exercise (SES)				

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (17 questions, range 0-68; higher scores indicate worse function); HOOS=Hip dysfunction and Osteoarthritis Outcome Score, (0-100; higher scores indicate better function/quality of life); AQoL-8D= Assessment of Quality of Life-8 Dimension (range -0.04 to 1.0; higher scores indicate better quality of life); IPEQ-W= Incidental and Planned Exercise Questionnaire subscore, planned sport activities, version W (range 0-42, higher scores represent higher levels of activity); ASES= Arthritis Self Efficacy Scale pain subscale (range 1-10, higher scores represent greater pain self-efficacy); BFMS= Brief Fear of Movement Scale for Osteoarthritis (range 6-24, higher scores represent more fear, increase indicates worse); SES= Self-efficacy for Exercise Scale (0-90; higher scores indicate greater self-efficacy).

[a]Data presented are complete cases and summarized as mean (standard deviation).

[b]Control: X n=x.

[c]Intervention: X n=x.

[d]Measured on an 11-point numeric rating scale for average hip pain severity during walking in the past week. Score range is 0 (no pain) to 10 (worst pain possible); higher score indicates worse pain.

[e]The Hip dysfunction and Osteoarthritis Outcome Score is a hip-specific questionnaire with different subscales, pain (10 items), function in sport and recreation (physical

function) (4 items) and hip-related quality of life (4 items). Total subscale scores range from 0 to 100, with 0 indicating extreme problems and 100, no problems.
[f]The Assessment of Quality of Life-8 Dimension is a 35-item questionnaire regarding health-related quality of life. The score range is -0.04 to 1.0; higher scores indicate better quality of life.

Table 3: Change in continuous outcomes within and between groups[a].

	Mean (SD) change within groups Baseline minus 24-weeks		Difference in change between groups Intervention vs. Control	
	Control (n=xx)[b]	Intervention (n=xx)[c]	Mean (95% CI)	P-value
Primary outcomes				
Hip pain severity during walking (NRS)[d]				
Physical function (WOMAC)[d]				
Secondary outcomes				
Hip dysfunction and Osteoarthritis Outcome Score[e]				
Pain				
Function in sport and recreation				
Hip-related quality of life				
Health-related quality of life (AQoL-8D)[e]				
Physical activity (IPEQ-W)[e]				
Fear of movement (BFMS)[d]				
Pain self-efficacy (ASES pain subscale)[e]				
Self-efficacy for exercise (SES)[e]				

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (17 questions, range 0-68; higher scores indicate worse function); HOOS=Hip dysfunction and Osteoarthritis Outcome Score, (0-100; higher scores indicate better function/quality of life); AQoL-8D= Assessment of Quality of Life-8 Dimension (range -0.04 to 1.0; higher scores indicate better quality of life); IPEQ-W= Incidental and Planned Exercise Questionnaire subscore, planned sport activities, version W (range 0-42, higher scores represent higher levels of activity); BFMS= Brief Fear of Movement Scale for Osteoarthritis (range 6-24, higher scores represent more fear); ASES= Arthritis Self Efficacy Scale pain subscale (range 1-10, higher scores represent greater pain self-efficacy); SES= Self-efficacy for Exercise Scale (0-90; higher scores indicate greater self-efficacy).

[a] Within group changes presented are complete cases and summarized as mean (standard deviation).

[b] Control: xxx n=x.

[c] Intervention: xxx n=x.

[d] For change within groups, positive changes indicate improvement. For difference in change between groups, positive differences favour Intervention.

[e] For change within groups, negative changes indicate improvement. For difference in change between groups, negative differences favour Intervention.

Table 4: Binary secondary outcomes.

	Control n/Total (%)	Intervention n/Total (%)	Relative risk (95% CI)[a]	P-value	Risk difference (95% CI)[a]	P-value
Achieved MCID in walking pain[b]						
Achieved MCID in function[c]						
Improved overall change in hip condition[d]						
Improved physical function[e]						
≥1 oral pain medication used[f]						

CI = confidence interval; MCID = minimum clinically important difference.

[a] Relative risks >1 and risk differences >0 favour xxx group.

[b] Participants achieving a pain reduction greater than or equal to the MCID of 1.8 NRS points classified as having achieved a clinically relevant pain reduction.

[c] Participants achieving an improvement in function greater than or equal to the MCID of 6 WOMAC points classified as having achieved a clinically relevant function improvement.

[d] Global improvement in overall change in hip condition rated using 7-point scale with terminal descriptors of 'much worse' to 'much better', with participants indicating 'moderately better' or 'much better' classified as improved and all others as not improved.

[e] Global improvement in physical function rated using 7-point scale with terminal descriptors of 'much less' to 'much more', with participants indicating 'moderately more' or 'much more' classified as improved and all others as not improved.

[f] Self-report use of at least 1 oral pain medication, defined as one or more of oral non-steroidal anti-inflammatory drugs and/or analgesics (paracetamol combinations) and/or oral glucocorticoids and/or oral opioids taken at least once a week in the prior month (4 weeks) for hip pain.

Table 5: Process measures.

	Control (n= xxx)	Intervention (n= xxx)
8 weeks		
Website visits, n (%)[a]		
Never		
1-5 times		
6-10 times		
>10 times		
Number of days home strengthening exercise performed, mean (SD)[b]	NA	
Use of the "My Exercise Messages" app, n (%)[c]		
Never	NA	
1-5 times	NA	
6-10 times	NA	
>10 times	NA	
16 weeks		
Website visits, n (%)[a]		
Never		
1-5 times		
6-10 times		
>10 times		
Number of days home strengthening exercise performed, mean (SD)[b]	NA	
Use of the "My Exercise Messages" app, n (%)[c]		
Never	NA	
1-5 times	NA	
6-10 times	NA	
>10 times	NA	
24 weeks		
Website visits, n (%)[a]		
Never		
1-5 times		
6-10 times		
>10 times		

Downloaded exercise programs, n (%)[d]		
Program 1	NA	
Program 2	NA	
Program 3	NA	
Number of days home strengthening exercise performed, mean (SD)[b]	NA	
Participants achieving 'acceptable adherence', n (%)[e]	NA	
Exercise Adherence Rating Scale [EARS] Section B, mean (SD)	NA	
Downloaded "My Exercise Messages" app, n (%)[f]	NA	
Frequency of entering exercise into the "My Exercise Messages" app, n (%)[g]		
Never	NA	
Rarely	NA	
Sometimes	NA	
Often	NA	
Always		
Use of the "My Exercise Messages" app, n (%)[c]		
Never	NA	
1-5 times	NA	
6-10 times	NA	
>10 times	NA	
Usefulness of the "My Hip Exercise" website, mean (SD)[h]	NA	
Usefulness of the "My Exercise Messages" app, mean (SD)[i]	NA	
Exercise equipment use, n (%)[j]	NA	
Registration to recommended external websites, n (%)[k]		
"Taking Control of Your Hip and Knee Osteoarthritis" online course	NA	
"PainTRAINER" online course	NA	
Leg strengthening exercise, n (%)[l]		NA

NA=not applicable; SD= Standard deviation; EARS= Exercise Adherence Rating Scale (Section B, 0-24, higher scores indicate better adherence).

[a] Self-reported. Participants asked "How many times did you visit the website over the past 2 months?".

[b] Self-reported. Participants asked "In the past two weeks, on how many days did you perform the strengthening exercises from the "My Hip Exercise" program? (0-14)" at 8, 16 and 24 weeks with the average number of days per week over the three timepoints reported.

[c] Self-reported. Participants asked "On how many days in the past 14 days did you use the "My Exercise Messages" app (including reading any notifications, entering data and/or opening the app)".

[d] Yes/No response to "Did you download any of the exercise programs from the website as PDF documents?" for programs 1, 2, and 3.
[e] 'Acceptable' adherence defined as performing the home strengthening exercise on an average of 2 days or more per week.
[f] Yes/No response to "Did you download the recommended app, My Exercises Messages?"
[g] Self-reported. Scored on a 5-point Likert scale for "How often did you enter your number of exercise days into the app as requested?"
[h] Self-reported. Scored on an 11-point NRS for "How useful did you find the "My Hip Exercise" website program? (not including the app)" from 0 = "not at all useful" to 10 = "extremely useful".
[i] Self-reported. Scored on an 11-point NRS for "How useful did you find the "My Exercise Messages" app?" from 0 = "not at all useful" to 10 = "extremely useful".
[j] Yes/No response to "Did you use exercise equipment (such as weights or elastic bands) to add resistance to your leg for your strengthening exercises?"
[k] Yes/No response to "Did you register for this course?" for 1) an online course "Taking Control of Your Hip and Knee Osteoarthritis," and 2) an online course "PainTRAINER"
[l] Yes/No response to "Over the past 24 weeks, did you perform leg strengthening exercise one or more times a week for at least 8 weeks?"

Table 6. Adverse events and co-interventions at 24 weeks.

	Control	Intervention
Adverse events, n (%)[a]	[n=xx]	[n=xx]
≥1 non-serious related adverse event		
Hip pain Pain in other areas		
Knee pain		
Ankle/foot pain		
Pain in other areas		
Hip stiffness		
Hip swelling		
Ceased trial participation due to non-serious related adverse event		
≥1 serious related adverse event[b]		
Current oral pain medication use, n (%)[c]	[n=xx]	[n=xx]
Non-steroidal anti-inflammatories		
Analgesia (paracetamol combinations)		
Oral corticosteroids		
Oral opioids		
Other co-interventions, n (%)[d]	[n=xx]	[n=xx]
≥1 co-intervention used		
Massage		
Manual therapy		
Walking stick/gait aid		
Transcutaneous Electrical Nerve Stimulation		
Injections		
Hydrotherapy		
Acupuncture		
Arthroscopic surgery		
Hip joint replacement surgery		

COX-2= cyclooxygenase-2.

NOTE- numbers of specific adverse events may exceed the number of participants reporting ANY event as it was possible for participants to report more than one type of adverse event. These adverse events can be considered as intercurrent events in this trial.

[a] Related adverse events are defined as "any problem experienced in the study hip or elsewhere in the body deemed to be a result of participating in the trial and at least one of i) caused negative/adverse symptoms/effects for two days or more, and/or ii) resulted in the participant seeking treatment or taking medication".

[b] Serious adverse events defined as any untoward medical occurrence that i) results in death; ii) is life-threatening; iii) requires hospitalisation or prolongation of existing inpatient hospitalisation; iv) results in persistent or significant disability or incapacity; v) is a congenital anomaly or birth defect, or; vi) any other important medical condition

which, although not included in the above, may require medical or surgical intervention to prevent one of the outcomes listed.
[c] Defined as used at least once per week over the past four weeks for hip pain.
[d] Self-reported use for hip pain over prior 24 weeks.

Table 7: Effects of the intervention on the primary outcomes at 24-weeks assuming acceptable adherence (with acceptable adherence defined as performing the home strengthening exercise on an average of 2 days or more per week).

	Difference in change (baseline minus 24-weeks) between groups[a] Intervention - Control Mean (95% CI)	**P-value**
Hip pain severity during walking (NRS)[b]		
Physical function (WOMAC)[c]		

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

[a] For difference in change between groups, positive differences favor Intervention.

[b] NRS, 0-10 with higher scores indicate more pain; MCID=1.8.

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

Figure 1. Flow of participants through the trial.