Protocol 'MappKO Protocol_Current clean_v4'
ACTRN12621000724875

Mobile app for Knee Osteoarthritis: the MappKO randomised controlled trial

Statistical Analysis Plan

Statistical Analysis Plan

TRIAL FULL TITLE	Evaluating the effects on exercise adherence and physical function of a Mobile app for Knee Osteoarthritis: the MappKO randomised controlled trial
TRIAL SHORT TITLE	MappKO
TRIAL REGISTRATION	ACTRN12621000724875
PROTOCOL VERSION	6 October 2022 (named 'MappKO Protocol_Current clean_v4')
TRIAL CHIEF INVESTIGATOR	Rana Hinman, University of Melbourne
TRIAL STATISTICIANS	Foong Mok, University of Melbourne Peixuan Li, University of Melbourne Karen Lamb, University of Melbourne Fiona McManus, University of Melbourne
SAP AUTHORS	Ximena Camacho, University of Melbourne Foong Mok, University of Melbourne Karen Lamb, University of Melbourne
SAP VERSION	1
SAP VERSION DATE	17-DEC-2024

SAP Revision History

Version	Reason(s) for change	Date
1.0	Initial version	17-DEC-2024

SAP Signatures

I give my approval for the attached SAP entitled MappKO dated 17/12/2024.

Chief Investigator

Name / Affiliation	Signature	Date
Rana Hinman, University of Melbourne	[REDACTED]	17/12/2024

SAP Author

Name / Affiliation	Signature	Date
Foong Mok, University of Melbourne	[REDACTED]	17/12/2024

Senior statistician

Name / Affiliation	Signature	Date
Karen Lamb, University of Melbourne	[REDACTED]	17/12/2024

Table of Contents

Abbreviations and Definitions

ABBREVIATION	DEFINITION
AE	Adverse event
CI	Confidence interval
OA	Osteoarthritis
SAP	Statistical analysis plan
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
KOOS	Knee Injury and Osteoarthritis Outcome Score
PASE	Physical Activity Scale for the Elderly
EARS	Exercise Adherence Rating Scale
uMARS	User version of the Mobile Application Rating Scale
GP	General practitioner
MAR	Missing at random
DSMB	Data Safety Monitoring Board
CONSORT	Consolidated Standards of Reporting Trials
cLDA	Constrained longitudinal data analysis

1 Introduction

1.1 Preface

There is no cure for osteoarthritis (OA). Clinical guidelines advocate exercise management as a core treatment for all people with knee OA (1). However, adherence to exercise is problematic (2), particularly after contact with a health professional has ceased. People with hip and/or knee OA are faced with many barriers to exercise participation and these have been mapped using behaviour change theory (3). Scalable interventions that target these behavioural domains may be effective in promoting long-term participation in exercise in people with OA.

Mobile apps are a promising intervention that may be helpful in supporting exercise. Apps are effective at supporting people to implement positive behavioural changes required for chronic disease management, including physical health (4). We developed a mobile app called "My Exercise Messages", adapted from our effective SMS intervention (5-7) that provided behaviour change tips to overcome exercise challenges and facilitate exercise participation. The app was developed to overcome limitations of the SMS program (which cannot be implemented at scale without some cost to the user and has limited flexibility in message content due to character count restrictions). To date, "My Exercise Messages" has not been evaluated for its effectiveness in the management of people with knee OA. This trial aims to provide the first evidence about the efficacy of the app in people with knee OA.

1.2 Purpose of the SAP

The purpose of this Statistical Analysis Plan (SAP) is to outline the pre-planned analyses to be completed to support the main publication of the MappKO study.

2 Study Objectives and Endpoints

2.1 Study Objectives

The primary objective is to determine if use of the "My Exercise Messages" app following a short course of physiotherapy for people with knee OA enhances longer-term benefits on (1) physical function and (2) exercise session adherence at 26 weeks post-randomisation, compared to a control group who only undergo a short course of physiotherapy without the app.

The primary hypothesis is that the addition of the "My Exercise Messages" app after a short course of physiotherapy will lead to greater improvements in physical function and better adherence to home exercise, compared to controls, at 26 weeks post-randomisation in people with knee OA.

The secondary objectives of this study are:

- To determine if the "My Exercise Messages" app enhances medium-term benefits on physical function and exercise session adherence (at 14 weeks), following a short course of physiotherapy for people with knee OA.
- To determine if the "My Exercise Messages" app confers additional benefit to a short course of physiotherapy on other clinical outcomes (severity of knee pain; function in sport and recreation; knee-related quality-of-life; global ratings of change; self-efficacy for exercise; physical activity levels; other measures of exercise adherence; satisfaction with treatment program) at 26 weeks.

- To determine if the "My Exercise Messages" app affects willingness to undergo knee replacement surgery.
- To explore potential moderators of treatment effect on physical function (the clinical co-primary outcome), based on the following *a priori* hypotheses:
    - Beliefs about exercise
        - The hypothesis is that participants who believe that exercise is less effective for managing knee OA at baseline will report less improvement in physical function with the My Exercise Messages app, compared to those who believe exercise is more effective.
    - Confidence using technology
        - The hypothesis is that participants who are less confident using technology will report less improvement in physical function with the My Exercise Messages app, compared to participants who are more confident.
- (Among the intervention group only) To determine engagement with, and usefulness of, the "My Exercise Messages" app using a range of process measures (app download; cessation of use; usage; usefulness; rating).

2.2 Endpoints

There are two primary endpoints: (change in) physical function and exercise session adherence at 26 weeks post-randomisation.

- Physical function
    - Measured using the physical function subscale score of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). This is extracted from the Knee Injury and Osteoarthritis Outcome Score (KOOS) where 17 questions ask about knee function, with Likert response options ranging from no dysfunction to extreme dysfunction. Total subscale score ranges from 0 (no dysfunction) to 68 (maximum dysfunction). Change in physical function score is calculated as the score at follow-up (26 weeks) minus baseline.
- Exercise session adherence
    - Number of days prescribed strength exercise sessions were performed over the previous fortnight at follow-up (26 weeks). Self-reported at follow-up via survey with options ranging from "0" to "6 or more". Responses will be converted to and analysed as a binary outcome for primary analysis, with the categories as "0 to 5" or "6 or more". In secondary analysis, this outcome will be considered as an ordinal outcome.

The secondary endpoints are:

- Medium-term benefits in physical function and exercise adherence:
    - Physical function: Change in physical function score between baseline and 14 weeks. Physical function scores defined as per primary endpoints above.
    - Exercise adherence: Exercise adherence (number of days on which prescribed strength exercise sessions were performed over the previous fortnight) at 14 weeks. Adherence defined as per primary endpoints above.
- Additional benefits to a short course of physiotherapy on other clinical outcomes:

- Severity of knee pain: Measured using self-reported severity of knee pain during walking in the past week on a 11-point numerical rating scale, from 0=no pain and 10=worst pain possible. Change is calculated as pain score at follow-up (26 weeks) minus baseline.
- Function in sport and recreation: Measured using KOOS sport and recreation subscale. Self-reported using 5 questions about function with sport and recreational activities in the last week, with Likert response options ranging from 'None' to 'Extreme'. Scores range from 0-100. Lower scores indicate worse function. Change is calculated as function score at follow-up (26 weeks) minus baseline.
- Knee-related quality of life: Measured using KOOS quality of life subscale. Measured using 4 questions about knee related quality of life experienced in the last week, with Likert response options for each. Scores range from 0-100. Lower scores indicate worse quality of life. Change is calculated as quality of life score at follow-up (26 weeks) minus baseline.
- Global ratings of change:
    - Knee pain: Scored using a 7-point global rating of change Likert scale with response options ranging from "much worse/less" to "much better/more" when compared to baseline. Participants indicating they are "moderately better/more" or "much better/more" at follow-up (26 weeks) will be classified as 'improved'. All other respondents will be classified as 'not improved'.
    - Knee function: Scored using a 7-point global rating of change Likert scale with response options ranging from "much worse/less" to "much better/more" when compared to baseline. Participants indicating they are "moderately better/more" or "much better/more" at follow-up (26 weeks) will be classified as 'improved'. All other respondents will be classified as 'not improved'.
    - Knee overall: Scored using a 7-point global rating of change Likert scale with response options ranging from "much worse/less" to "much better/more" when compared to baseline. Participants indicating they are "moderately better/more" or "much better/more" at follow-up (26 weeks) will be classified as 'improved'. All other respondents will be classified as 'not improved'.
- Self-efficacy for exercise: Measured using Self-efficacy for Exercise scale, a nine-item scale that assesses self-efficacy expectations about ability to continue exercising in the face of perceived barriers. Total scores from 0 to 90, higher scores indicating higher self-efficacy. Change is calculated as self-efficacy score at follow-up (26 weeks) minus baseline.
- Physical activity levels: Measured using Physical Activity Scale for the Elderly (PASE). Measured using 10 questions about frequency and duration of recreational, household and occupational physical activity undertaken over the past 7 days. Scores range from 0 to 793; higher scores indicate greater levels of physical activity. Change is calculated as physical activity score at follow-up (26 weeks) minus baseline.
- Exercise adherence (analysed as an ordinal outcome): Number of days prescribed strength exercise sessions were performed over the previous fortnight at follow-up (26 weeks). Self-reported at follow-up via survey with options ranging from "0" to "6 or more".

  - Other measures of exercise adherence: Measured using Exercise Adherence Rating Scale (EARS). Measured using the 6-item section B of the EARS. Scores range from 0 to 24, with higher score indicating better adherence. Change is calculated as exercise adherence score at follow-up (26 weeks) minus baseline.
  - Satisfaction with treatment program: Measured using a 7-point global rating of change scale with response options from “extremely unsatisfied” to “extremely satisfied”. Participants indicating they are “moderately satisfied” or “extremely satisfied” at follow-up (26 weeks) will be classified as ‘satisfied’. Note: treatment program includes both the exercise program and app where applicable.
- Willingness to undergo knee replacement surgery: Measured using self-report. Participants will be asked about their willingness to have surgery on their knee in the near future, ranging from “definitely not willing” to “definitely willing” on a 5-point ordinal scale. Those who answered “probably willing” or “definitely willing” will be classified as ‘willing to undergo surgery’; those answering “unsure”, “probably unwilling”, or “definitely unwilling” will be classified as ‘unsure/not willing’. Participants will be asked about their willingness to undergo surgery at baseline and follow-up (26 weeks); at each time, the n (%) will be calculated.
- Moderators of treatment effect:
  - Beliefs about exercise for knee OA: Measured using self-report. Participants will be asked “How effective do you believe exercise is for managing knee osteoarthritis?” with Likert response options of “not at all effective”, “somewhat effective”, “moderately effective” and “highly effective”. Participants will be asked about their beliefs at baseline and follow-up (26 weeks); at each time, the n (%) selecting each response option will be reported. Responses will be dichotomised into ‘less effective’ (“not at all effective” and “somewhat effective”) and ‘more effective’ (“moderately effective” and “highly effective”). The belief at baseline will be used as a moderator in the analysis of the primary endpoints, considered in categories of ‘less effective’ and ‘more effective’.
  - Confidence using technology: Measured using self-report. Participants will be asked “How confident are you in using technology in everyday life?” with 4-point Likert response options “not at all confident”, “somewhat confident”, “moderately confident” and “extremely confident”. Participants will be asked about their beliefs at baseline and follow-up (26 weeks); at each time, the n (%) selecting each response option will be reported. Responses will be dichotomised into ‘less confident’ (“not at all confident” and “somewhat confident”) and ‘more confident’ (“moderately confident” and “extremely confident”). The confidence at baseline will be used as a moderator in the analysis of the primary endpoints, considered in categories of ‘less confident’ and ‘more confident’.

The following secondary endpoints (process measures) will be assessed among the intervention group only and will explore engagement with and usefulness of the “My Exercise Messages” app. All survey questions were mandatory and no missing responses are expected:

- Download of the “My Exercise Messages” app: Measured using email survey conducted at follow-up (26 weeks). Participants will indicate if they downloaded the app (yes/no), from which app store (Apple App Store/Google Play), and on which devices (smartphone/tablet). The n (%) selecting each response option will be reported.

- Number of participants who stopped using the "My Exercise Messages" app: Measured using email survey conducted at follow-up (26 weeks). Participants will indicate whether they stopped using the app (yes/no). Participants responding "yes" will be asked to specify when and why. The n (%) selecting each response (yes/no) will be reported; reasons and timing for stopping the app will be reported descriptively.
- Use of the "My Exercise Messages" app: Measured using email survey conducted monthly following randomisation. Participants will indicate if they have used the used the app in the past 14 days (yes/no). Participants responding "yes" will indicate the frequency of use (response options of: 1-3 days; 4-6 days; 7-10 days; 11-13 days; every day). The n (%) selecting each response will be reported.
- Usefulness of the "My Exercise Messages" app: measured using email survey conducted at follow-up (26 weeks). Participants will be asked to indicate their agreement on 8 questions with 7-point Likert response options ranging from 'strongly disagree' to 'strongly agree':
    i) "The My Exercise Messages app helped me manage my knee problems."
    ii) "The My Exercise Messages app reminded/prompted me to exercise"
    iii) "The My Exercise Messages app helped me monitor/track how often I performed my exercise"
    iv) "The messages I received from the My Exercise Messages helped me stick to my exercise program"
    v) "The messages I received from the My Exercise Messages app helped me overcome any difficulties I experienced in doing my exercise"
    vi) "The messages I received from the My Exercise Messages app helped motivate me to do my exercises"
    vii) "The number of messages I received from the My Exercise Messages app was the right amount for me."

  The n (%) of participants selecting each response option will be reported.
- Rating of the "My Exercise Messages" app: Measured using the user version of the Mobile Application Rating Scale (uMARS) sections A, B, D, F at follow-up (26 weeks).
    - Engagement: measured using subscale score for engagement (5 items). Engagement score will be calculated by averaging scores across items within the domain (i.e., across the 5 items).
    - Functionality: measured using subscale score for functionality (4 items). Functionality score will be calculated by averaging scores across items within the domain (i.e., across the 4 items).
    - Information quality: measured using subscale score for information quality (4 items). Information quality score will be calculated by averaging scores across items within the domain (i.e., across the 4 items).
    - Perceived impact: reported as individual items measured on a Likert scale with responses ranging from 1 to 5 points (1="strongly agree," 5="strongly disagree").

3 Study Methods

3.1 General Study Design and Plan

This project is a pragmatic, two-arm parallel randomised controlled superiority trial. Recruitment occurred across all Australian states/territories. Eligible participants are those who meet the National Institute for Health and Care Excellence clinical criteria for OA, have experienced at least 3

months of knee pain, experience knee pain on most days, and report at least mild physical dysfunction. All participants will receive exercise-based care by a physiotherapist. This will involve two telehealth consultations with a physiotherapist (using Zoom videoconferencing) over two weeks, for prescription of an individualised strengthening exercise program. All participants will receive resistance bands for exercising. After completing the physiotherapy consultations, participants will be encouraged to continue with their home exercises for the next 24 weeks. Following physiotherapy, around half of participants will be randomly allocated to download and engage with a mobile app (on their smartphone or tablet) for the 24-week home exercise period. Outcomes will be collected by web-based survey at baseline and 26 weeks after randomisation. In addition, the two primary outcomes will also be measured by web-based survey at 14 weeks after randomisation.

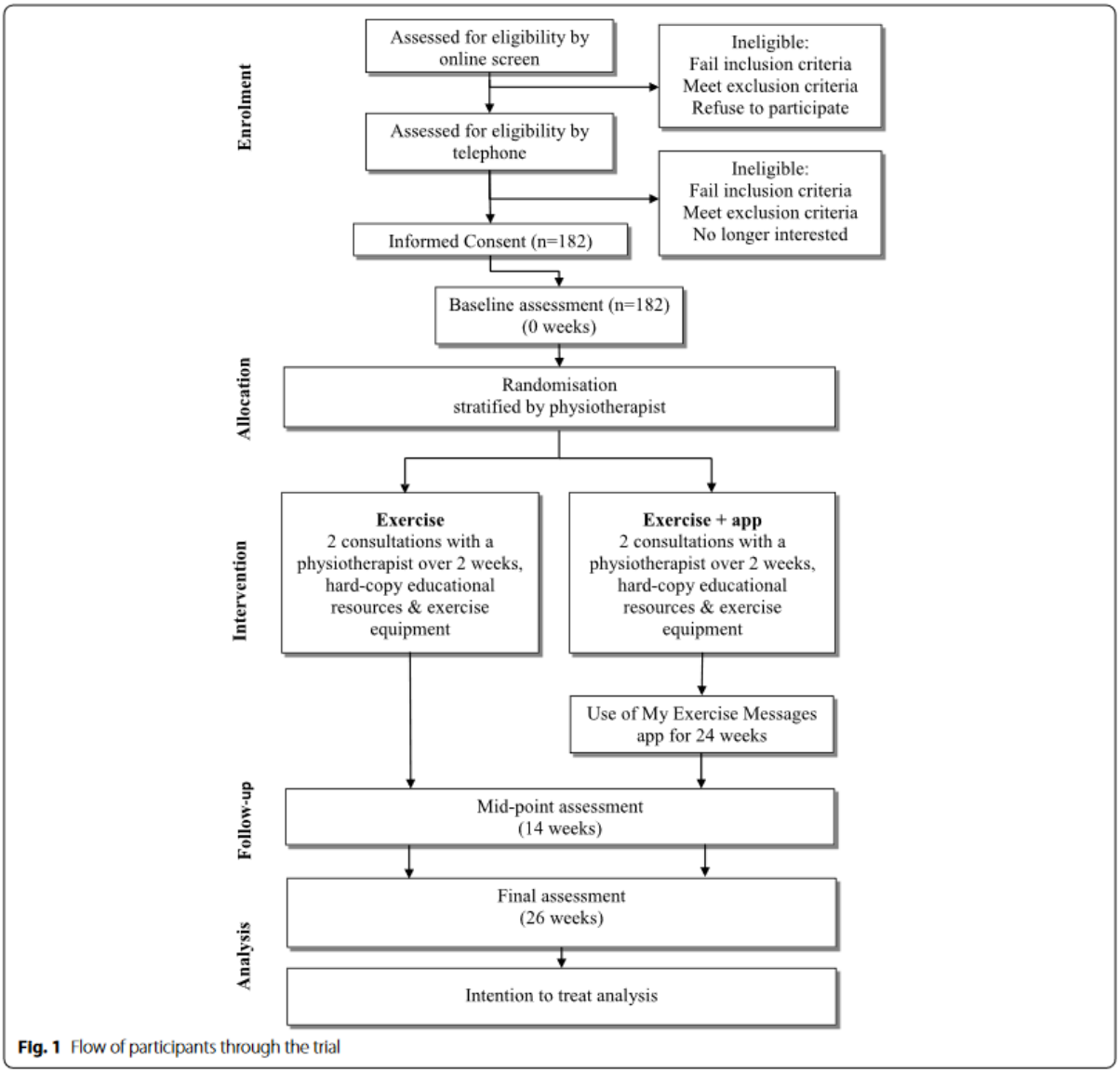

Fig. 1 Flow of participants through the trial

Figure 1. Flowchart showing participant timeline.

3.2 Equivalence or Non-Inferiority Studies

N/A

3.3 Inclusion-Exclusion Criteria and General Study Population

Eligibility criteria are:

- Meeting National Institute for Health and Care Excellence (8) clinical criteria for OA:
  - age ≥ 45 years;
  - activity-related knee joint pain;
  - morning knee stiffness ≤ 30 minutes.
- Reporting history of knee pain ≥ 3 months;
- Reporting knee pain on most days of the past month;
- Reporting at least mild physical dysfunction (score >20 out of 68 on WOMAC physical function subscale);
- Having access to a computer/laptop/tablet/smartphone with internet connection for Zoom consultations;
- Owning a smartphone with software compatible with the app (for using app if allocated to intervention group);
- Willingness to participate in video consultations for physiotherapy appointments; and
- Willingness to download and engage with an app regularly if allocated to intervention group.

Participants will be ineligible for the study if they meet any of the following exclusion criteria:

- inability to speak or read English;
- on waiting list for/planning knee/hip surgery in next 6 months;
- previous arthroplasty on affected knee;
- recent knee surgery (past 6 months);
- consulting/ed physiotherapy or doing regular (at least once per week) strengthening exercise for knee (past 6 months);
- self-reported inflammatory arthritis (e.g., rheumatoid arthritis);
- any neurological condition affecting lower limbs;
- any unstable/uncontrolled cardiovascular condition;
- history of fall (past 12 months) and no General Practitioner (GP) clearance to participate;
- house-bound due to immobility and no GP clearance to participate; and/or
- fail the Exercise and Sports Science Australia stage 1 pre-exercise screening questions (9) (consists of seven questions which are answered YES or NO, and are designed to identify individuals with signs or symptoms of underlying disease, or who may be at higher risk of an adverse event during exercise) and no GP clearance to participate.

3.4 Randomisation and Blinding

Participants will be randomly assigned to intervention ("My Exercise Messages" app) or control using a randomisation schedule prepared by a biostatistician (permuted random block sizes), stratified by physiotherapist. The schedule was stored on a password-protected website (REDCap) at the University of Melbourne maintained and accessed by researchers not involved in either participant recruitment or administration of primary/secondary outcome measures.

Treating physiotherapists and trial statisticians will be blinded to participants' assignment to the intervention. Participants allocated to the My Exercise Messages intervention group will not be told of their group allocation until after they have attended their two physiotherapy sessions, in order to ensure they do not disclose their group allocation to their physiotherapist (and therefore unblind the blinded therapists).

The trial statisticians will remain blinded until the database has been cleaned, a blinded data review has taken place, and the database is ready for analysis.

3.5 Study Variables

Physical function:

Physical dysfunction is measured via the WOMAC Osteoarthritis Index (Likert version 3.1). It is a valid, reliable and responsive self-reported OA-specific tool. The physical function subscale has 17 questions which evaluate knee function over the prior week during a variety of daily activities (response options ranging from 'no dysfunction' to 'extreme dysfunction'). It is mandatory to answer all questions. Total score ranges from 0 to 68 and higher scores indicate poorer function. Change scores will be determined with data from baseline and 26 weeks (or baseline and 14 weeks for secondary outcomes).

Exercise adherence (number of days strengthening exercises were performed in previous fortnight):

At 14 and 26 weeks, participants are asked "In the past two weeks, how many days did you perform the strengthening exercise program your physiotherapist prescribed for your knee problem?" Options range from "0" to "6 or more". This outcome will be dichotomised to "6 or more" (no/yes) for primary analysis. This outcome indicates complete adherence to prescribed exercise. In secondary analysis, the ordinal outcome will be examined.

Knee pain during walking:

Average pain on walking (over the prior week) is measured with an 11-point numerical rating scale that has terminal descriptors of 'no pain' (score = 0) and 'worst pain possible' (score = 10). Change scores will be calculated using data from baseline and 26 weeks.

Sport and recreation function:

The KOOS sport and recreational activities sub-scale measures function in sports and recreational activities. It comprises five questions ascertaining function with sport and recreational activities over the last week, with 5-point Likert response options ranging from 'none' to 'extreme'. The five possible answer options are scored from 0 (No problems) to 5 (Extreme problems) and the total

score is calculated as the sum of the items included. The score is transformed to a 0-100 scale representing the percentage of total possible score achievable. Lower scores indicate more difficulty with sports and recreational activities. Change scores will be calculated using data from baseline and 26 weeks.

Knee-related quality of life:

The KOOS quality of life subscale measures knee-related quality of life. It asks four questions about knee-related quality of life over the previous week, with 5-point Likert response options. The five possible answer options are scored from 0 (No problems) to 5 (Extreme problems) and the total score is calculated as the sum of the items included. The score is transformed to a 0-100 scale representing the percentage of total possible score. Lower scores indicating poorer quality of life. Change scores will be calculated using data from baseline and 26 weeks.

Physical activity:

Physical activity over the prior week is measured via the PASE. It comprises 10 questions about frequency and duration of recreational, household and occupational physical activity over the prior 7 days. Each activity is given a weighting, and the total score is computed by multiplying the duration spent in each activity by the respective weight, and summing over all activities. Total scores range from 0 to 793, with higher scores indicative of more physical activity. Change scores will be calculated using data from baseline and 26 weeks.

Exercise self-efficacy:

The Self-Efficacy for Exercise Scale measures self-efficacy expectations about ability to continue exercising in the face of perceived barriers. This 9-item tool has scores which range from 0 to 90, with higher scores indicating better exercise self-efficacy. Change scores will be calculated using data from baseline and 26 weeks.

Participant-perceived global changes:

At 26 weeks, participants rate their global change from baseline in i) knee pain, ii) physical function, and iii) their knee overall, via individual 7-point Likert scales (ranging from “much worse/less” to “much better/more”). On each scale, participants who are moderately better or much better will be classed as improved and the rest as not improved.

Satisfaction with exercise program (including the mobile app, if applicable):

At 26 weeks, satisfaction with the entire exercise program (including the mobile app for those who were randomised to receive it) is rated with a 7-point Likert scale (ranging from “extremely unsatisfied” to “extremely satisfied”). Participants who are moderately satisfied or extremely satisfied will be classed as satisfied and the rest as not satisfied.

Exercise Adherence Rating Scale (EARS):

A secondary measure of exercise adherence is included (in addition to the primary adherence measure). At 26 weeks, participants rate their adherence to their prescribed home strengthening exercises using Section B of the EARS. This has six items and each is scored with a 5-point scale (terminal descriptors of 'strongly agree' to 'strongly disagree'). The total score ranges from 0 to 24, with higher scores indicating better adherence.

Willingness to undergo knee joint replacement:

At baseline and 26 weeks, participants will answer the question "How willing are you to have a knee joint replacement on your knee in the near future?" using a 5-point Likert scale (ranging from "definitely not willing" to "definitely willing"). Participants will be dichotomised into those not willing (definitely not willing; probably not willing) or willing/unsure (unsure; probably willing; definitely willing).

Beliefs about exercise:

At baseline and 26 weeks, all participants will answer the question "How effective do you believe exercise is for managing knee osteoarthritis?" using a 4-point Likert scale (with response options of "not at all effective"; "somewhat effective"; "moderately effective"; and "highly effective").

Process measures – intervention group (exercise + app) only:

A range of self-reported process measures will be collected from the exercise plus app group only at 26 weeks. These include: i) downloads of the 'My Exercise Messages' app; ii) number of people who stop using the app (and reasons why); iii) number of days the app was used in the prior 14 days (via monthly survey until 26 weeks); iv) usefulness of the app; and v) user engagement, functionality, information quality and perceived impact of the app using the user Mobile Application Rating Scale.

4 Sample Size

There are two primary outcomes in this study. The minimal clinically important difference on the clinical primary outcome physical function was 6 units on WOMAC at 26 weeks. Assuming a between-participant standard deviation of 12 units (7), a baseline to 26-week correlation of 0.5 (7), an ANCOVA adjusted for baseline score, and an adjusted alpha of 0.025 to account for the two primary outcomes, we needed 58 participants per arm to achieve 80% power. To detect a difference of 1.2 exercise sessions (in the prior 2 weeks) in the adherence primary outcome at 26-week follow-up between groups (based on data from a previous trial evaluating SMS behaviour change messages (6)), assuming a standard deviation of 2.4, 80% power and alpha of 0.025, 77 participants per arm were required. This assumed that the number of exercise sessions was collected as a continuous variable. As physiotherapists treated participants in both arms of the trial, we did not adjust the sample size for clustering by physiotherapist. Allowing for 15% attrition, we aimed to recruit 91 people/arm, with a total sample of 182 participants, to have sufficient power for both primary outcomes.

The target sample size remained 182 since the trial had commenced recruitment when the decision was made to collect the adherence primary outcome (number of exercise sessions in the previous fortnight) as ordinal data.

5 General Considerations

5.1 Timing of Analyses

Final analysis will be performed after all (n=182) participants have reached the 26-week endpoint and been provided the opportunity to complete outcome measures, and after the Statistical Analysis Plan has been finalised and published on our Centre's website.

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 investigators and study biostatisticians prior to database lock (defined as such as the state is kept constant, i.e., it is locked so that data cannot be subsequently amended). During this 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 breaking the mask along with the reason for classifying dyad to "as randomised" and "as treated".
3. Participant's overall adherence to study intervention, in relation to the use of the participant's data (or part of it). Overall adherence will be listed and tabulated for the purpose of the blinded data review meeting.
4. Participant's inclusion or exclusion status with regards to each analysis population.
5. 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 random 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, random 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 database and any changes to this SAP after unmasking will be documented in an amendment of this SAP and considered post-hoc analyses.

5.2 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:

- A participant receiving a different intervention to what they were allocated to
- Adverse events
- Pain medication
- Knee surgery
- Use of a different exercise app
- Subsequently report a comorbidity that should have deemed them ineligible to participate

Intercurrent events will be summarised by each intervention group.

5.3 Analysis Populations

The following analysis populations are planned. The status of each participant with regards to the populations and “as randomised”/”as treated” treatment group will be finalised during the blinded data review meeting.

5.3.1 Full Analysis Population

This will consist of all participants who were randomised, excluding participants who have withdrawn from the study and indicated they would like all their existing data collected to withdrawal date to be removed from the database, as documented in the minutes of the blinded data review meeting. Participants will be reported and analysed according to the randomised study arm (“as randomised”). This population will be used in the analysis of the effectiveness study variables. This will be the intention-to-treat population.

5.3.2 Safety Population

This will consist of all participants who received at least one study treatment (including those who only received the physiotherapy treatment but not the app). Receipt of at least one study treatment is defined as attending *at least one physiotherapy consult* for both arms. Participants will be reported and analysed according to the intervention received. This population will be used in the analysis of the safety study variables.

5.4 Covariates and Subgroups

5.4.1 Stratifying covariates

The models used for primary and secondary endpoints will be adjusted for the stratification variable (treating physiotherapist).

5.4.2 Confounding covariates

No sensitivity analysis will be undertaken to account for potential confounding variables (i.e., variables known to be prognostic of the primary outcome which could have chance imbalance between arms).

5.4.3 Subgroups

To assess whether the effect of the mobile app on the clinical primary outcome (physical function) is moderated by either of the pre-specified moderators (beliefs about exercise and confidence using technology, described in Section 2.2), exploratory subgroup analyses will be performed for the change in physical function at 26 weeks. The subgroups are:

1) Beliefs about the effectiveness of evidence for managing knee osteoarthritis
2) Confidence using technology

Subgroup responses will be grouped into binary categories as described in Section 2.2. The subgroup (main effect) and its interaction with the intervention arm will be added to the unadjusted model to evaluate whether the intervention effect ("My Exercise Messages" app + physiotherapy versus physiotherapy alone) differs between subgroup categories. Results of the subgroup analyses will be displayed using forest plots, presenting the estimate and two-sided 95% confidence interval (CI) of the treatment effect within each subgroup level and including the *p*-value for heterogeneity of the treatment effect between the subgroups.

5.5 Missing Data

To describe the missing data, the frequency and percentage of participants with a missing value at baseline, medium-term (14 weeks), and follow-up (26 weeks), will be summarised for the primary outcome variables overall and by intervention arm for the intention-to-treat population. Baseline and demographic characteristics will be summarised overall and for those with and without a missing value at each of the three time points separately to examine if any characteristics appear to be associated with the presence or absence of data.

As the primary strategy to handle missing data, the analysis of continuous study variables will use a likelihood-based approach to deal with missing data (i.e., constrained longitudinal data analysis). This approach assumes that the probability of missing data on the study variable is not related to the missing data but to some of the observed measured data in the model. That is, it assumes that the data are Missing at Random (MAR).

In addition, missing primary outcome data (where the outcome has greater than 5% missing data) will be multiply imputed using chained equations. The imputation model will include age, gender, height, weight, body mass index, employment status, co-morbidities and medications and it will be performed separately by treatment group. The number of imputed data sets will be at least equal to the percentage of missing data in the available case analysis, with a minimum number of 20 imputations used if the percentage of missing data is greater than 5% but less than 20%.

5.6 Interim Analyses and Data Monitoring

No interim analysis was conducted. As this was a low-risk trial, no data safety monitoring board (DSMB) was required. Adverse events were self-reported by participants at follow-up (26 weeks).

5.7 Multiple Testing

This study has two primary outcomes: physical function and number of exercise sessions completed. To account for the dual outcomes, the alpha level will be adjusted to 0.025.

There are numerous secondary outcomes, which are all exploratory and not powered for. There will therefore not be any adjustment for multiple secondary outcomes; instead, the estimates, 95% CIs, and *p*-values will be reported in order to let readers use their own judgment about the relative

weight of the conclusions regarding the intervention group effects on secondary outcomes. This approach aligns with the usage of *p*-values favoured by the American Statistical Association (10).

6 Summary of study data

6.1 Subject disposition

The flow of participants will be presented in a Consolidated Standards of Reporting Trials (CONSORT) diagram (11). The following will be reported:

- Number of participants assessed for eligibility
- Number of participants excluded due to not meeting the inclusion criteria, declined to participate and any other reasons
- Number of participants allocated to each of the two intervention arms
- Number of participants at follow-up and number of participants who withdrew from the study and who were lost to follow-up
- Number of participants included in the analysis of each primary outcome

6.2 Derived variables

All outcomes are derived variables (See Section 3.5).

6.3 Protocol deviations

No protocol deviations were specified in the study protocol. The major protocol deviations that could occur are being allocated the incorrect intervention (i.e., being told to download the app despite being randomised to the control group) or failing to download the app (i.e., not receiving the treatment). 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 the 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 were assigned to (i.e., intervention arm if incorrectly instructed to download the app but in the control group). Sensitivity analysis will be performed only for the primary outcomes.

6.4 Demographic and baseline variables

Demographic and baseline variables for participants will be summarised and presented by each intervention arm and overall using frequencies and percentages for categorical variables, mean and standard deviation for continuous variables, or median and quartiles (25^{th} and 75^{th} percentile) for non-symmetrical continuous variables. If variables have missing values, we will report the non-missing sample either in the table or in the footnote of the table. No *p*-values will be reported to compare the two arms.

6.5 Concurrent illnesses and medical conditions

Participants will report at baseline any concurrent illnesses and medical conditions, and will report any adverse events at 26 weeks. The information will be collated in the descriptive statistics table for baseline data, and adverse events table for data at 26 weeks.

6.6 Prior and concurrent medications

Participants will report at baseline and follow-up (26 weeks) if they were receiving any pain and arthritis medications and co-interventions, and whether they accessed any medical / healthcare mobile apps for their knee pain. This information will be collated in the descriptive characteristics

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 any other co-intervention once in the past month will be reported as a recent user. This will be summarised by intervention / no 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 arms.

6.7 Treatment adherence

In this study, all participants were advised by their physiotherapist to perform home strengthening exercises 3 times/week. A primary outcome in this trial is the number of days strengthening exercises were performed (over the previous fortnight), which was measured and will be reported as described in Section 2.2. A secondary outcome is the Exercise Adherence Rating Scale, which was measured and will be reported as described in Section 2.2. Additional measures of participant adherence are shown below in the Table and data from these measures will be reported using descriptive statistics as appropriate for each treatment group (a sample table is included in the appendix):

Adherence measure	Description	Scale
Attendance at physiotherapist consultations (collected in control and intervention groups)	Recorded by physiotherapist in treatment notes for each consultation	Number of consultations attended
Download of the "My Exercise Messages" app (intervention group only)	Participants reported (Yes/No) via survey at 26 weeks if they downloaded the app, from which app store, and on which devices.	Number (%) of participants reporting Yes and No will be reported. Number (%) of smartphone and tablets, and App Store and Google Play, will be reported.
Number of participants who stopped using the "My Exercise Messages" app (intervention group only)	Participants indicated via survey at 26 weeks whether they stopped using the app (Yes or No) and if so, when and why.	Number (%) of participants reporting Yes and No will be reported. When and why will be reported descriptively.
Use of the "My Exercise Messages" app (intervention group only)	Participants self-reported (once every month) usage of the app in response to the following question: "In the last 14 days, did you use the My Exercise Messages app?" If yes, on how many days (in the last 14 days) did you use the My Exercise Messages	Number (%) of participants will be reported for each response option each month.

	app? Response options of: 1-3 days; 4-6 days; 7-10 days; 11-13 days; every day.	

7 Effectiveness analyses

7.1 Primary effectiveness analysis

The primary estimand for MappKO is defined according to the addendum to the ICH E9 on estimands and sensitivity analysis in clinical trials. MappKO aims to answer the specific research questions: does the addition of the "My Exercise Messages" app following a short course of physiotherapy, compared to the short course of physiotherapy alone, improve i) physical function of people with knee OA at 26-weeks, as measured by the mean difference in WOMAC, and ii) adherence to exercise , as measured by the number of days strengthening exercises were performed (6+ days vs <6 days), regardless of the post-randomisation (intercurrent) events listed below.

Primary estimand attributes

- Treatment: "My Exercise Messages" app + short course of physiotherapy compared to short course of physiotherapy alone (control), allocation by randomisation.
- Population: Adults who meet the National Institute for Health and Care Excellence clinical criteria for OA, report history of knee pain over at least 3 months, report knee pain on most days of the past month, report at least mild physical dysfunction who have access to Zoom consultations, own a smartphone and are willing to participate in video consultations and to engage with an app regularly.
- Variables: i) WOMAC at 26 weeks post-randomisation; ii) adherence to exercise at 26 weeks post-randomisation.
- Intercurrent events: Possible intercurrent events outlined in Section 5.2 will be handled using the treatment policy strategy.
- Population level summaries: i) mean difference in WOMAC between arms ("My Exercise Messages" app + short course of physiotherapy vs. short course of physiotherapy alone) and ii) risk difference and risk ratio in adherence between arms ("My Exercise Messages" app + short course of physiotherapy vs. short course of physiotherapy alone).

The number and percentage of participants who had intercurrent events will be described overall and by treatment arm.

The primary outcome – WOMAC – will be analysed using a constrained longitudinal data analysis (cLDA) model, with response consisting of all scores (baseline, 14 weeks, and 26 weeks) and the model including factors representing intervention (physiotherapy alone [the reference group], physiotherapy + "My Exercise Messages" app), time and intervention by time interaction with the restriction of a common baseline mean score across interventions. This refers to the assumption that at baseline there are no differences between the interventions in the mean score, thus assuming the randomisation was effective. This assumption will be enforced in the statistical model. The model will include random effects for participant and the stratification factor, physiotherapist.

The primary outcome – adherence – will be analysed using a logistic regression model fit using generalized estimating equations to account for physiotherapist as the clustering factor. The response

will consist of the binary outcome at 26 weeks (primary time point), and the model will include a factor representing intervention (physiotherapy alone [the reference group], physiotherapy + "My Exercise Messages" app) and clustering by physiotherapist. The primary analysis will be undertaken on multiply imputed data should more than 5% of participants have missing data on this outcome

The primary hypotheses will be evaluated using the estimate, multiplicity adjusted two-sided 95% confidence interval, and *p*-value of the i) mean difference in WOMAC between the two treatment arms in the change from baseline to 26 weeks post-randomisation and ii) risk difference and risk ratio for adherence at 26 weeks post-randomisation.

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 effectiveness analyses

The following secondary outcomes will be treated as continuous outcomes and will be analysed similarly to the primary outcome, although captured only at baseline and 26 weeks, and targeting a similar primary estimand: i) Change in severity of knee pain during walking; ii) Change in KOOS sport and recreation subscale; iii) Change in KOOS quality of life subscale; iv) Change in Physical Activity Scale for the Elderly; v) Change in Self-Efficacy for Exercise Scale. Exercise Adherence Rating Scale is only measured at 26 weeks and will be compared between treatment arms using a linear-mixed model with a factor representing intervention and a random effect for physiotherapist. Appropriate transformations may be applied to the variables before fitting the model if considered skewed.

Binary outcomes will be analysed using logistic regression fitted using generalised estimating equations to account for clustering of participants within physiotherapist: i) Adherence at 14 weeks; ii) Global rating of change in knee pain at 26 weeks; iii) Global rating of change in knee function at 26 weeks; iv) Global rating of change in knee overall at 26 weeks; v) Satisfaction with the exercise program at 26 weeks; vi) Willingness to undergo knee replacement surgery at 26 weeks. The model for willingness to undergo knee replacement surgery at 26 weeks will adjust for the baseline measure of this outcome. All other binary outcomes listed were only captured at 26 weeks.

Finally, the adherence primary outcome at 26 weeks will be analysed as an ordinal outcome in secondary analyses. Ordinal logistic regression models will be fitted to adherence at 26 weeks, accounting for clustering by physiotherapist. The proportional odds assumption will be assessed using the Brant test. If proportional odds cannot be assumed, multinomial regression models will be fitted. Results will be reported as odds ratios, or relative risk ratios if multinomial regression is used, with corresponding 95% CIs.

7.3 Exploratory effectiveness 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. Minimum adherence to the intervention will be defined as having reported 'Yes' to the question "In the last 14 days, did you use the My Exercise Messages app?" in Month 1 and 'at least 4-6 days' in the subsequent question about frequency of use asked of those who had used the app. In a supplementary analysis, the estimate corresponding to a secondary estimand will be obtained that will apply to those participants who would adhere under any of the interventions rather than all randomised participants. A supplementary analysis will estimate the treatment effect on the two primary outcomes assuming adherence to the intervention, as defined in Section 6.5. A Complier Average Causal Effect analysis including all randomised dyads will be undertaken to estimate the average effect of "My Exercise Messages" + physiotherapy compared to physiotherapy only on each outcome for participants who would adhere to whichever intervention group they were assigned to, considering a binary definition of adherence. The Complier Average Causal Effect will be estimated using a latent class regression model on the multiple imputed data if multiple imputation is required (Section 5.5), with two latent classes for compliers and non-compliers (15). Complier Average Causal Effects will be reported with 95% CIs and *p*-values. The number and percentage of dyads who adhered to the intervention will be described.

7.4 Safety analyses

In this trial, adverse events were self-reported by participants at 26 weeks, and defined as "any problem experienced in the study knee or elsewhere in the body deemed by the participant to be a result of participating in the trial AND at least one of i) that caused increased pain and/or disability for two days or more, and/or ii) resulted in the participant seeking treatment from a health professional".

The following will be reported:

- the number (and proportion) of participants reporting related adverse events for each treatment group as outlined in the safety population and the nature of the event(s) described;
- 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 adverse events and their types (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).

8 Reporting conventions

The standard deviation will be reported with mean values, and interquartile range (25^{th} and 75^{th} percentiles) with the median. Percentages will be reported without decimal points, and percentages less than 1% will be presented as <1%. All values with a decimal point will be replaced with a middle dot (·). *P*-values will be reported to three significant figures. and capped at three decimal places. *P*-values less than 0.001 will be reported as "<0·001". In any cells in a row of data for which no data are reported, 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.

9 Technical details

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

10 Summary of changes to the protocol

This statistical analysis plan outlines the details of the analyses beyond what was specified in the protocol.

- Adherence Primary Outcome
  The adherence primary outcome was captured as an ordinal outcome instead of the originally planned continuous outcome. It was decided that this should be treated as a binary outcome in primary analysis and an ordinal outcome in sensitivity analysis. Methods to analyse these types of outcome have been described in the SAP.

- 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 was outlined in the SAP.

11 References

1. Victoria AO. A problem worth solving : the rising cost of musculoskeletal conditions in Australia : a report. Arthritis & Osteoporosis Victoria; 2013.
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014;73(7):1323-30.
3. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis. 2013;72(7):1125-35.
4. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22(3):363-88.
5. (UK) NIfHaCE. Osteoarthritis: Care and Management in Adults: National Institute for Health and Care Excellence (UK); 2014.
6. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;1(1):CD004376.
7. Uthman OA, van der Windt DA, Jordan JL, Dziedzic KS, Healey EL, Peat GM, Foster NE. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. BMJ. 2013;347:f5555.
8. Lim BW, Hinman RS, Wrigley TV, Sharma L, Bennell KL. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. Arthritis Rheum. 2008;59(7):943-51.
9. Pisters MF, Veenhof C, Schellevis FG, Twisk JW, Dekker J, De Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. Arthritis Care Res (Hoboken). 2010;62(8):1087-94.
10. Hinman RS, Hunt MA, Simic M, Bennell KL. Exercise, Gait Retraining, Footwear and Insoles for Knee Osteoarthritis. Current Physical Medicine and Rehabilitation Reports. 2013;1(1):21-8.

11. Bennell KL, Kyriakides M, Metcalf B, Egerton T, Wrigley TV, Hodges PW, et al. Neuromuscular Versus Quadriceps Strengthening Exercise in Patients With Medial Knee Osteoarthritis and Varus Malalignment. Arthritis Rheumatol. 2014;66(4):950-9.

Appendix A – Table Listing for the Final Report

Heading	Title	Population Set
Table 1	Baseline characteristics of participants, by intervention group reported as mean (standard deviation) unless otherwise stated.	Randomised
Table 2	Mean (SD) scores on continuous outcome measures across time, by treatment group.	Randomised
Table 3	Change from baseline within groups on continuous outcomes across time	Randomised
Table 4	Binary outcomes, and changes in willingness to undergo knee replacement surgery at 13 and 26 weeks, along with adjusted relative risks and risk differences. Counts and proportions based on complete case data.	Randomised
Table 5	Adherence and process measures.	Randomised
Table 6	Adverse events and co-interventions for knee pain reported at 26 weeks, according to group, presented as number (%) of participants.	Safety
Table 7	Moderation of the treatment effect on the primary outcome of physical function[a] at 26 weeks by pre-specified moderators.	Randomised

Appendix B – Templates for tables

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

	Control group (n=xxx)	Intervention group (n=xxx)
Age (years)		
Gender, n (%)		
Female		
Male		
Non-binary		
Prefer not to say		
Other		
Height (m)		
Weight (kg)		
Body mass index (kg/m^2)		
Ethnicity, n (%)		
Australian/New Zealand		
Aboriginal and Torres Strait Islander		
European		
Asian		
Other Oceania		
North African & Middle Eastern		
Sub-Saharan Africa		
North American		
South American		
Other		
Geographical location, n (%)[a]		
Major cities		
Inner regional		
Outer regional		
Remote		
Very remote		
Symptom duration (years)		

Education level, n (%)

<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

Retired (not due to health reasons)

Unemployed/student/homemaker

Unable to work due to health reasons

Comorbidities, n (%)

Heart disease

High blood pressure

Lung disease

Diabetes

Ulcer or stomach disease

Kidney disease

Liver disease

Anaemia or other blood disease

Cancer

Depression

Osteoarthritis

Back pain

Rheumatoid arthritis

Other medical problem/s

Trial treatment expectations, n (%)[b]

No effect

Minimal improvement

Moderate improvement

Large improvement

Complete recovery

More confident using technology in day to day life, n (%)[c]

Use of technology, n (%)[d]

Searching internet for health information

Apps to manage health

Social media (e.g. Facebook) for health information/support

Wearable devices/trackers for managing health

YouTube for health information

Subscription to online health management programs

Oral medications for knee, n (%)[e]

Analgesics (paracetamol combinations)

Opioid analgesics

Non-steroidal anti-inflammatories

COX-2 inhibitors

Corticosteroids

Belief about exercise for knee osteoarthritis, n (%)

Not at all effective

Somewhat effective

Moderately effective

Highly effective

Willing to undergo knee replacement surgery, n (%)[f]

---

[a]based on residential postcode, in accordance with Australian Statistical Geography Standard.

[b]recorded prior to randomisation using a 5-point Likert scale with anchors of "no effect at all" to "complete recovery".

[c]rated using a 4-point Likert scale with options of "not at all confident", "somewhat confident", "moderately confident", and "extremely confident". Dichotomised into 'less confident' ("not at all confident" and "somewhat confident") and 'more confident' ("moderately confident" and "extremely confident").

[d]usage over previous month, each rated as yes or no.

[e]used at least once per week over the past month for knee pain.

[f]rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing", with those indicating "probably willing" or "definitely willing" classified as 'willing', and all other options classified as 'unwilling'.

COX-2= cyclooxygenase-2.

Note: If data are missing for a variable then an additional row will be included reporting the number of missing observations for that variable

Table 2. Mean (SD) scores on continuous outcome measures across time, by treatment group.

	Baseline		14 weeks		26 weeks	
	Control (n=xxx)	**Intervention (n=xxx)**	**Control (n=xxx)**	**Intervention (n=xx)**	**Control (n=xx)**	**Intervention (n=xx)**
Primary outcome						
Physical function (WOMAC)						
Secondary outcomes						
Knee pain during walking (NRS)			NA	NA		
Sport & recreation function (KOOS)			NA	NA		
Knee-related quality of life (KOOS)			NA	NA		
Physical activity (PASE)			NA	NA		
Self-efficacy for exercise (SES)			NA	NA		
Exercise adherence (EARS)	NA	NA	NA	NA		

WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function); NA=not applicable; NRS=numerical rating scale (0-10; higher scores indicate worse pain); KOOS=Knee Injury and Osteoarthritis Outcome Score, (0-100; higher scores indicate better function/quality of life); PASE=Physical Activity Scale for the Elderly (0->400; higher scores indicate better physical activity); SES= Self-efficacy for Exercise Scale (0-90; higher scores indicate greater self-efficacy); EARS= Exercise Adherence Rating Scale (Section B, 0-24, higher scores indicate better adherence).

Table 3: Change from baseline within groups on continuous outcomes across time from baseline, and difference in change between groups.

	Mean (SD) change within groups				Difference in change between groups			
	Baseline minus week 14		Baseline minus week 26		Baseline to week 14		Baseline to Week 26	
	Control (n=xx)	Intervention (n=xx)	Control (n=xx)	Intervention (n=xx)	Mean difference (95% CI)	P-value	Mean difference (95% CI)	P-value
Primary outcomes								
Physical function (WOMAC)[a]								
Secondary outcomes								
Knee pain during walking (NRS)[a]	NA	NA			NA			
Sport & recreation function (KOOS)[b]	NA	NA			NA			
Knee-related quality of life (KOOS)[b]	NA	NA			NA			
Physical activity (PASE)[b]	NA	NA			NA			
Self-efficacy for exercise (ASES)[b]	NA	NA			NA			
Exercise adherence (EARS)[c]	NA	NA			NA			

WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function); NA= not applicable; NRS=numerical rating scale (0-10; higher scores indicate worse pain); KOOS=Knee Injury and Osteoarthritis Outcome Score, (0-100; higher scores indicate better function/quality of life); PASE=Physical Activity Scale for the Elderly (0->400; higher scores indicate better physical activity); SES= Self-efficacy for Exercise Scale (0-90; higher scores indicate greater self-efficacy); EARS= Exercise Adherence Rating Scale (Section B, 0-24, higher scores indicate better adherence).

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

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

[c]Only measured at 26 weeks so no change calculated from baseline. For difference between groups, positive differences favour Intervention.

Table 4: Binary outcomes, and changes in willingness to undergo knee replacement surgery at 14 and 26 weeks, along with adjusted relative risks and risk differences. Counts and proportions based on complete case data.

	Control (n=xx)	Intervention (n=xx)	Relative Risk[a] (95% CI)	P- value	Risk Difference[a] (95% CI)	P-value
Primary outcomes						
Exercise adherence (binary; 26 weeks)						
Secondary outcomes						
Exercise adherence (binary; 14 weeks)						
Improved pain[b]						
Improved function[b]						
Improved knee overall[b]						
Satisfied with exercise program[c]						
Change in willingness for surgery[d]						
Remain unwilling						
Unwilling to willing						
Willing to unwilling						
Remain willing						

[a]Risk differences > 0 and relative risks > 1 favour Control group

[b]Rated using 7-point scale with terminal descriptors of "much worse/less" to "much better/more", with those indicating "moderately better/more" or "much better/more" classified as 'improved'.

[c]Rated using 7-point scale with terminal descriptors of "extremely unsatisfied" to "extremely satisfied", with those indicating "moderately satisfied" or "extremely satisfied" classified as 'satisfied', noting that participants allocated to the intervention group were asked to consider the mobile app as part of their exercise program.

[d]Rated using a 5-point scale with terminal descriptors of "definitely not willing" to "definitely willing", with those indicating "probably willing" or "definitely willing" classified as 'willing to have knee surgery in the near future', and all other options classified as 'unwilling'.

Table 5: Adherence and process measures.

	Control (n= xxx)	Intervention (n=xxx)
Median (IQR) physiotherapy consults attended[a]		
Attended zero consults, n (%)		
Attended one consult, n (%)		
Attended two consults, n (%)		
Belief about exercise for knee osteoarthritis, n (%)[b]		
Not at all effective		
Somewhat effective		
Moderately effective		
Highly effective		
Use of My Exercise Messages app:	NA	
Downloaded the app, n (%)[c]		
Downloaded the app from App Store, n(%)		
Downloaded the app from Google Play, n(%)		
Used app on smart phone, n(%)		
Used app on tablet, n(%)		
Stopped using the app, n (%)[c]		
App use over past 14 days, n (%)[d]		
Month 1:		
None		
1-3 days		
4-6 days		
7-10 days		
11-13 days		
Every day		
Month 2:		
None		
1-3 days		
4-6 days		
7-10 days		
11-13 days		
Every day		
Month 3:		
None		
1-3 days		
4-6 days		
7-10 days		
11-13 days		
Every day		
Month 4:		
None		
1-3 days		
4-6 days		
7-10 days		
11-13 days		
Every day		
Month 5:		
None		
1-3 days		
4-6 days		

7-10 days	
11-13 days	
Every day	
Month 6:	
None	
1-3 days	
4-6 days	
7-10 days	
11-13 days	
Every day	
Usefulness of the app, n (%):[e]	NA
Helped me manage my knee problems	
Reminded/prompted me to exercise	
Helped me monitor/track how often I exercised	
Helped me stick to my exercise program	
Helped me overcome any difficulties with exercise	
Helped motivate me to do my exercises	
Number of messages I received was the right amount for me	
Mobile App Rating Scale scores:	NA
App engagement, mean (SD)	
App functionality, mean (SD)	
App information quality, mean (SD)	
Perceived Impact:	
Awareness (app increased my awareness of importance of exercise for knee problems)	
1 (strongly disagree), n (%)	
2	
3	
4	
5 (strongly agree)	
Knowledge (app increased my knowledge/understanding of exercise)	
1 (strongly disagree), n (%)	
2	
3	
4	
5 (strongly agree)	
Attitudes (app increased my knowledge/understanding of exercise)	
1 (strongly disagree), n (%)	
2	
3	
4	
5 (strongly agree)	
Intention to change (app increased my intentions/motivation to exercise)	
1 (strongly disagree), n (%)	
2	
3	
4	
5 (strongly agree)	
Help seeking (app would encourage me to seek further help to exercise if I needed it)	
1 (strongly disagree), n (%)	
2	
3	

4
5 (strongly agree)

[a]as recorded by physiotherapists in treatment notes.

[b]measured at 26 weeks.

[c]as reported (yes/no) by participants in survey at 26 weeks.

[d]self-reported by participants (once every month) in response to the question: "In the last 14 days, did you use the My Exercise Messages app?" "If yes, on how many days (in the last 14 days) did you use the My Exercise Messages app?" Response options were: 1-3 days; 4-6 days; 7-10 days; 11-13 days; every day.

[e]rated at 26 weeks via a 7-point Likert scale with response options ranging from "strongly disagree" to "strongly agree", with those indicating "moderately agree" or "strongly agree" classified as 'agree' and all others as 'disagree'.

NA=not applicable

Table 6. Adverse events and co-interventions for knee pain reported at 26 weeks, according to group, presented as number (%) of participants.

	Control (n=xx)	Intervention (n=xx)
Adverse events:		
N reporting ANY adverse event		
Event description		
Event description		
Event description		
Event description		
Event description		
Event description		
Ceased trial participation as result		
Serious adverse events[a]		
Co-intervention use:[b]		
Oral analgesics (paracetamol combinations)		
Oral opioid analgesics		
Oral non-steroidal anti-inflammatories		
Oral COX-2 inhibitors		
Oral corticosteroids		
Topical anti-inflammatories		
Injections		
Land-based exercise other than that prescribed in RCT		
Heat/cold treatment		
Massage		
Knee braces		
Manual therapy		
Orthotics/arch supports		
Walking stick/gait aid		
Hydrotherapy		
Electrotherapy[c]		
Acupuncture		
Health/medical mobile apps		
Arthroscopy		
Knee replacement surgery		

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]Serious adverse events defined as any untoward medical occurrence that resulted in death, was life-threatening, required hospitalization or resulted in significant disability.

[b]self-reported use over prior month. For oral medications, participants must have used a medication at least once a week in the past month to be considered a user of that medication.

[C]includes ultrasound, transcutaneous electrical nerve stimulation and low level laser therapy.

COX-2= cyclooxygenase-2

Table 7: Moderation of the treatment effect on the primary outcome of physical function[a] at 26 weeks by pre-specified moderators.

Moderator	Mean (SD) change within Control (n=xx)	Mean (SD) change within Intervention (n=xx)	Intervention -Control (95% CI)	Interaction P-value
Beliefs about exercise:				
Less effective				
More effective				
Confidence using technology:[b]				
Less confident				
More confident				

[a]measured using the Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function).

[b]rated using a 4-point Likert scale with options of not at all confident, somewhat confident, moderately confidant, and extremely confident. Dichotomised into less confident (not at all and somewhat confident) and more confident (moderately and extremely confident).

Appendix C – Appendix Tables

Appendix Table 1: Ordinal adherence outcome at 14 and 26 weeks. Counts and proportions based on complete case data.

	Control (n=xx)	Intervention (n=xx)	Relative Risk[a] (95% CI)	P- value	Risk Difference[a] (95% CI)	P-value
Secondary outcomes						
Exercise adherence (ordinal; 26 weeks)						
0						
1						
2						
3						
4						
5						
6 or more						
Exercise adherence (ordinal; 14 weeks)						
0						
1						
2						
3						
4						
5						
6 or more						

Appendix Table 2: Additional analyses on the primary outcomes at 26 weeks.

	Mean difference[b] (95% CI)	P-value
Physical function (WOMAC)[a]		

	Relative Risk[c] (95% CI)	Risk Difference[c] (95% CI)	P-value
Exercise adherence (binary)			

[a]WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function)

[b]For difference in change between groups, positive differences favour Intervention.

[c]Risk differences > 0 and relative risks > 1 favour Control group