Statistical Analysis Plan (SAP)

Trial: Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. The PEAK randomised controlled trial.

Contents

Section 1. Administrative Information

1. Title

Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis (OA). The PEAK randomised controlled trial.

2. Trial registration

Prospectively registered (Australian New Zealand Clinical Trials Registry Trial Id: ACTRN12619001240134, 09/09/2019)

3. SAP version

Version: 1.0 Date: 05/05/2023

4. Protocol Version

This document has been written based on information contained in the PEAK study protocol version 5 dated 8/10/2021. A manuscript outlining the protocol was published as follows:

Hinman, R.S., Kimp, A.J., Campbell, P.K. *et al.* Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. Protocol for the PEAK randomised controlled trial. *BMC Musculoskelet Disord* **21,** 522 (2020). https://doi.org/10.1186/s12891-020-03523-8

5. SAP Revisions

Not applicable

6. Names and affiliations

Document prepared by Professor Rana Hinman, Mrs Penny Campbell, Mr Alexander Kimp, Professor Kim Bennell, University of Melbourne; Professor Trevor Russell, Professor Nadine Foster, University of Queensland; and Professor Anthony Harris, A/Prof Jessica Kasza, Monash University.

Emails: RS Hinman ranash@unimelb.edu.au; PK Campbell penelope.campbell@unimelb.edu.au; AJ Kimp alexander.kimp@unimelb.edu.au; KL Bennell k.bennell@unimelb.edu.au; NE Foster n.foster@uq.edu.au; T Russell t.russell@uq.edu.au; A Harris anthony.harris@buseco.monash.edu.au; J Kasza jessica.kasza@monash.edu.

Signatures:

Signature of senior statistician responsible (A/Prof Jessica Kasza): *Date: 05/05/2023*

Signature of chief investigator (Prof Rana Hinman): *Date: 05/05/2023*

Section 2: Introduction

7. Brief background and rationale

Exercise is a core recommended strategy for management of knee osteoarthritis (OA). Physiotherapists are frequent providers of exercise care to people with knee OA and care is traditionally provided face-to-face in clinical settings. However, video-conferencing can provide telehealth-delivered care to people located geographically remotely from the clinician. This can increase access to physiotherapy care and reduce burden on patients related to time and costs of travel to the clinic. It remains unknown if outcomes from care provided via video-conferencing are non-inferior to face-to-face care.

Refer to the published protocol for a detailed background to this trial:

Hinman, R.S., Kimp, A.J., Campbell, P.K. *et al.* Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. Protocol for the PEAK randomised controlled trial. *BMC Musculoskelet Disord* **21,** 522 (2020). https://doi.org/10.1186/s12891-020-03523-8

8. Objectives

Research hypothesis:

We hypothesise that video consultations with a physiotherapist lead to non-inferior knee pain on walking and/or physical function outcomes at 3 months, compared to face-to-face consultations in people with knee OA.

Study objective:

Primary objective: To determine whether video consultations with a physiotherapist lead to non-inferior outcomes (knee pain on walking and/or physical function) at 3 months compared to face-to-face consultations.

Secondary objectives:

i) To determine whether outcomes of care from video consultations are non-inferior to outcomes from face-to-face consultations with respect to improving knee pain on walking and physical function at 9 months.
ii) To compare clinical outcomes of care between video and face-to-face consultations on other measures (health-related quality of life; therapeutic relationship; global ratings of change; satisfaction with care; self-efficacy; physical activity levels) at 3 and 9 months.
iii) To determine if attending video consultations is associated with reduced participant-level time and financial costs, as well as greater convenience, compared to face-to-face consultations, at 3 months.
iv) To compare direct and indirect healthcare usage (related to knee pain and trial participation) across video and face-to-face consultations at 9 months.
v) To explore potential moderators of treatment effect on the primary outcomes and on participant time and travel (secondary outcomes), based on the following a priori hypotheses:
    a. Experience with online video platforms at baseline
        Hypothesis- Participants who are less frequent users of video platforms will have less improvement in primary outcomes with video consultations (relative to face-to-face), compared to participants who are more frequent users.
    b. Geographical residence
        Hypothesis- Participants who don't live in major city areas will report reduced participant-level time and travel distance (secondary outcomes) with video consultations (relative to face-to-face), compared to participants who live in major cities.
    c. Beliefs about physiotherapy care delivery at baseline
        Hypothesis- Participants who believe that video consultations are less effective for managing musculoskeletal problems will report less improvement in primary outcomes with video

consultations (relative to face-to-face), compared to those who believe video consultations are more effective.

d. Confidence using technology at baseline

Hypothesis- Participants who are less confident with using technology will report less improvement in primary outcomes with video consultations (relative to face-to-face), compared to participants who are more confident.

Section 3: Trial Methods

9. Trial design

Pragmatic comparative effectiveness, two parallel arm, non-inferiority randomized controlled trial (RCT).

10. Randomisation

The randomisation schedule was prepared by the biostatistician (permuted block sizes 6 to 12) stratified by physiotherapist. Where physiotherapists treated participants at two clinic locations, randomisation was stratified by location within each physiotherapist. The schedule was stored on a password-protected website (REDCap) at Uni of Melbourne maintained by a researcher not involved in either participant recruitment or administration of primary/secondary outcome measures. Group allocation was revealed by this same researcher after baseline assessment was completed.

11. Sample size

Sample size was based on detecting non-inferiority of video consultations relative to face-to-face at 3 months after randomisation. For change in numerical rating scale (NRS) pain, a non-inferiority margin (NIM) of 0.95 units was chosen as this is less than the lowest of the range (1.0-2.0 units)[1, 2] reported as the minimum clinically important difference (MCID) by people with chronic pain, and less than the MCID of 1.75 units (extrapolated from a 100 mm visual analogue scale (VAS))[3] for OA by clinician consensus. For Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) subscales, a NIM of 8 mm on VAS versions of WOMAC (score 0-100) is used in drug non-inferiority RCTs,[4] as it is less than the MCIDs of 9.1-9.3 mm.[5, 6] We used the Likert version of WOMAC (scored 0-68) for function, thus our chosen NIM was 5.44 units (extrapolated from 8 mm). Assuming standard deviations (SD) of changes from baseline of 2.8 and 15 units for pain and function respectively and correlations of 0.3 between baseline and follow-up,[7, 8] 15% loss to follow-up, 90% power, and a one-sided 2.5% significance level, we need 197 people/arm for pain and 172/arm for function, a total of 394 people.

12. Framework

This trial uses a non-inferiority hypothesis testing framework between groups for the primary outcomes. A superiority hypothesis testing framework will be used for secondary outcomes. Results will be reported in accordance with the non-inferiority trials extension of the CONSORT 2010 statement.[9]

13. Statistical interim analyses and stopping guidance

Nil

14. Timing of final analysis

Final analysis will be performed after all (n=394) participants have reached the 9-month timepoint and been provided the opportunity to complete outcome measure and after the Statistical Analysis Plan has been finalised and published on our Centre's website.

15. Timing of outcome assessments

Outcomes are collected at baseline, 3 months, 6 months (health care usage data only) and 9 months.

Section 4: Statistical Principles

16. Level of statistical significance

All applicable statistical tests will be 2-sided and will be performed using a 5% significance level, including those conducted within the non-inferiority framework.

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

We have two primary outcomes: knee pain and physical function. We will not adjust for dual outcomes but instead report all effect sizes, confidence intervals, and p values in order to let readers use their own judgment about the relative weight of the conclusions on the comparison between videoconferencing and face-to-face care outcomes for knee OA. This approach aligns with the usage of p-values favoured by the American Statistical Association.[10]

18. Confidence intervals to be reported

All confidence intervals will be two-sided 95% confidence intervals.

19. Adherence and Protocol Deviations

In this study, participants are asked to attend 5 consultations with the physiotherapist over 3 months. Participants will be classified as 'adherent' to the intervention (or not) based on the number of consultations attended (as documented in physiotherapist consultation notes). Randomised participants who attended ≥3 consultations will be classified as 'adherent' to the intervention. The number (%) of adherent and non-adherent participants will be reported for each treatment group.

20. Analysis Populations

Except where stated otherwise, all analyses will be conducted under the principle of intention-to-treat, whereby all participants are included in their randomised groups. Previously, in the published trial protocol, we had stated that non-inferiority would be assessed using the per-protocol datasets (i.e. restricting the analysis to only those participants who attended ≥3 consultations). However, in light of the fact that the intention-to-treat effect is likely to be of greater interest than a per-protocol effect, and the difficulties in interpretation of the per-protocol effect, we now plan to conduct the primary analyses using the intention-to-treat sample, in line with contemporary recommendations for non-inferiority trials.[11, 12]

Section 5: Trial Population

21. Screening Data

Screening data will be collected and summarised. A CONSORT flow diagram will be used.[13] The following summaries will be presented in text and/or flow diagram: time frame for recruitment, the number of participants screened, the number of participants recruited, the number of screened participants not recruited, and the reasons for non-recruitment.

22. Eligibility

Trial inclusion criteria are described in the trial protocol. Reasons for exclusion will be summarised in the CONSORT [13] flow diagram.

Consented,

23. Recruitment

A CONSORT flow diagram [13] will be used to describe the number of people consented to participate, randomised, allocated to each treatment group, those lost to follow up (including reasons) and analysed.

24. Withdrawal/follow-up

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

25. Baseline descriptive measures

Baseline characteristics will be summarised by treatment group and presented in a table (see Appendix for all planned tables):

- Recruitment source
- Data collection mode
- Height, weight, body mass index (BMI)
- Age
- Gender
- Duration of knee OA symptoms
- Geographical location
- Education level
- Current employment status
- Problems in other joints
- Comorbidities
- Expectation of treatment outcome
- Confidence using technology in day to day life
- Frequency of use of technology
- Beliefs about effectiveness of different modes of physiotherapy care delivery
- Prior experience with different modes of physiotherapy care delivery

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

imbalance will be noted.

An appendix table will provide summaries of baseline characteristics and baseline levels of primary and secondary outcomes (as relevant, as not all secondary outcomes could be measured at baseline) and compare these characteristics between two groups: those participants who provide both primary outcomes at 3 months, and those participants who are missing one or both primary outcomes at this time-point. T-tests will be used to compare continuous characteristics between these groups, and chi-squared tests will be used to compare categorical characteristics.

Section 6: Analysis

26. Outcome definitions

Name	Description	Scale	Outcome definition
Primary Outcomes			
Severity of knee pain during walking[14]	Scored on an 11-point NRS for average pain on walking in the last week.	Ranges from 0 to 10; where 0=no pain and 10=worst pain possible.	Change score at 3 months (primary time-point) and 9 months (secondary time-point) will be calculated as follow-up minus baseline.
Physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[46]	Scored using 17 questions regarding knee function in the last 48 hours, with Likert response options ranging from no dysfunction to extreme dysfunction.	Ranges from 0 (no dysfunction) to 68 (maximum dysfunction).	Change score at 3 months (primary time-point) and 9 months (secondary time-point) will be calculated as follow-up minus baseline.
Secondary Outcomes			
Quality of life (AQoL-6D)[15]	Scored using the 20-item Assessment of Quality of Life II Instrument (6D version), which covers the topics of Independent Living, Relationships, Mental Health, Coping, Pain and Senses to come up with one overall value representing quality of life.	Total score ranges from −0.04 to 1.00; higher scores indicate better quality of life.	Change score at 3 months and 9 months will be calculated as follow-up minus baseline.
Physical Activity scale for the elderly (PASE)[16]	Scored via 10 questions about frequency and duration of recreational, household and occupational physical activity undertaken over the past 7 days.	Scores range from 0 to 400+ ; higher scores indicate greater levels of physical activity	Change score at 3 months and 9 months will be calculated as follow-up minus baseline.
Arthritis Self Efficacy Scale[17]	Using the 8-item Arthritis Self-Efficacy Scale. Participants rate their ability to do 8 tasks from 1 (very uncertain) to 10 (very certain).	Total scores are an average of the 8 items with a range from 1 to 10; higher scores indicate higher self-efficacy.	Change score at 3 months and 9 months will be calculated as follow-up minus baseline.
Global rating of change in: a) Knee pain b) Physical function c) Physical activity	Scored at 3 and 9 months only, using a 7-point global rating of change Likert scale.	Response options range from “much worse/less” to “much better/more” when compared to baseline.	At 3 and 9 months, those indicating they are “moderately better/more” or “much better/more” will be classified as improved. All others will be classified as not improved. N (%) will be reported for each group.
Satisfaction with the physiotherapy consultations	Scored at 3 and 9 months only, using an 11-point NRS, for “How satisfied are you with the physiotherapy consultations you received in this study?”	Response options from “0=extremely unsatisfied” to “10=extremely satisfied”.	Score at 3 and 9 months.
Working Alliance Inventory Short Form[18]	Scored separately by both the patient (at 3 months)	Overall scores range from 12 to 84 (with higher scores	Score at 3 months (or final consultation)

	and the physiotherapist (after the 5th consultation or on the day the 5th consultation was due to be scheduled for participants who cancel/do not attend), based on 12 statements relating to the perceived trust and agreement between the therapist and client, each rated on a 7-point scale.	indicating a stronger therapeutic alliance).	
Convenience of consultations	Scored at 3 months only on an 11-point NRS for "Overall, how would you rate the convenience of your consultations with the physiotherapist?"	Ranges from 0 to 10; 0=extremely inconvenient and 10= extremely convenient	Score at 3 months
Attendance at consultations	Recorded by physiotherapists for each consultation throughout the 3 month physiotherapist intervention period.	Number of consultations attended (ranges from 0 to 5 maximum)	Number of consultations at 3 months
Adherence with strengthening exercise program.	Rated at 3 and 9 months only on a 11-point NRS for "I have been doing my exercises exactly as I was asked to by my PEAK trial physiotherapist (number of sessions, exercises and repetitions)"	Ranges from 0 to 10; where 0= strongly disagree and 10= strongly agree.	Adherence score at 3 and 9 months
Strengthening exercise sessions performed over the past week.	Self-reported as how many days over the prior week that the strength exercises were performed, at 3 and 9 months only.	Range from 0 to 7	Number of sessions at 3 and 9 months
Adherence with physical activity plan.	Rated at 3 and 9 months only on a 11-point NRS for "I followed the physical activity plan that my PEAK trial physiotherapist helped me to develop"	Ranges from 0 to 10; 0= strongly disagree and 10= strongly agree.	Adherence score at 3 and 9 months
Adverse events	Reported by participants using survey open-ended questions at 3 months and 9 months.	Defined as "any problem experienced in the study knee or elsewhere in the body deemed by the participant to be a result of the exercises, physical activity plans and/or advice given by the physiotherapist AND at least one of i) that caused increased pain and/or interfered with function for two days or more, and/or ii) resulted in the participant seeking treatment from a health professional"	At 3 and 9 months, the number (%) of participants reporting ANY adverse event for each group. The nature of specific events (including serious adverse events, defined as any untoward medical occurrence that resulted in death, was life-threatening, required hospitalization or resulted in significant disability) reported will be described along with the number (%) of participants reporting each. The number (%) of

			participants who discontinue the intervention because of an adverse event will be reported.
Co-intervention use	Obtained via a custom-developed survey to indicate the type and frequency of visits to health care providers (excluding the physiotherapy consults delivered as part of the RCT), purchase of prescription and over the counter medication, injections, hospitalisation and use of investigative procedures for their knee pain only (and/or as a result of trial participation in the case of follow-up surveys). Retrospective recall over 3 month intervals. Collected at baseline, 3, 6 & 9 months.	Any participant who has "used" any co-intervention at least once will be classed as a user of that intervention, and all others as a non-user.	At baseline, 3 and 9 months the n (%) users of: a) Purchased any oral/topical medication for their knee pain; b) Consulted any health professional for their knee pain (excluding physiotherapy consults as part of the RCT). The mean number of visits to health professionals will also be reported.
Participant time	Participants will record total time spent per consultation (travelling to/from, waiting and consultation time) in a log-book.	Recorded in minutes.	For each participant, at 3 months. the: a) Time for initial consultation b) Average time across all follow-up consultations attended will be reported.
Physiotherapist time	Therapists will record the total time spent in each consultation (excluding note-taking and scheduling)	Recorded in minutes.	For each participant, at 3 months, the a) Time for initial consultation b) Average time across all follow-up consultations attended will be reported.
Participant travel	Distance travelled to and from appointment location, and mode of transport to consultations (including need for another person to accompany) will be recorded for each consultation attended by participants in a log-book.	Total distance travelled recorded in kms. For each participant, the most common mode of transport across all consultations attended will be determined as either car, public transport, walking, no dominant mode evident, no transport required.	At 3 months, the mean distance travelled for consultations for each participant will be reported. N (%) for each transport mode will be reported for each group, along with n (%) of participants requiring an accompanying person for at least one consultation.

27. Analysis methods

Between-group differences in mean change in pain and function at 3 and 9 months (baseline minus follow-up) will be compared using mixed linear regression models including terms for baseline, randomised group, time,

and an interaction between time and randomised group, and the stratifying variable of physiotherapist. Random effects will be included for participants. If a physiotherapist treats participants at two locations, two separate terms for that physiotherapist, corresponding to each location, will be included. Non-inferiority will be demonstrated if the lower bound of the two-sided 95% CI for between-group difference (video minus face-to-face) is above -0.95 for pain and/or -5.44 for function, at 3 months (the primary time-point), however we will also test for non-inferiority on the primary outcomes at the secondary time-point of 9 months. 95% CIs correspond to testing the null hypothesis of non-inferiority at a one-sided significance level of 2.5%. Due to difficulties with their interpretation, p-values associated with non-inferiority hypotheses are not commonly reported and will not be reported here.[19]

Mixed linear regression models as described above will be fit to compare the continuous outcomes of AQoL-6D, PASE and ASES. Mixed linear regression models as above but without adjustment for baseline values will be fit for adherence with strengthening exercise program, number of exercise sessions performed, and adherence with physical activity plan. Linear regression models including terms for randomised group and physiotherapist will be fit for working alliance inventory short form (separate models will be fit for scores provided by the participants and physiotherapists), satisfaction, convenience of consultations, attendance at consultations, the participant and physiotherapist time variables, and participant travel distance. Global ratings of change will be analysed using log-binomial regression models, including terms for randomised group and physiotherapist, fit via generalised estimating equations with an exchangeable working correlation structure to account for the multiple measurements per participant and robust standard error estimation. Results will be presented as relative risks and 95% confidence intervals. Binomial regression models with an identity link function will also be fit in order to present risk differences and 95% confidence intervals. Assumptions will be assessed using standard diagnostic plots and methods. Since these are all secondary analyses, p-values for these assessments will not be adjusted for multiple comparisons.

If non-inferiority of a primary outcome is demonstrated, the superiority of the outcome will then be assessed and declared if the lower bound of the two-sided 95% CI for between-group difference exceeds zero. Since these are all secondary analyses, p-values for these assessments will not be adjusted for multiple comparisons.[20, 21]

Since we expect that most participants in the video consultations arm will not be required to travel to their appointments and not require an accompanying person, the outcomes of mode of transportation and requiring an accompanying person will be presented descriptively by treatment group as counts and percentages.

We will interpret and report findings transparently and separately for each primary outcome. For example, if we observe no difference in function between groups but significantly increased pain with video consultations, we will conclude that video consultations are non-inferior to face-to-face care for improving function but are inferior for pain relief, in knee OA. Patients seeking care, and physiotherapists delivering care, will then be fully informed about benefits, and limitations, of video consultations compared to face-to-face care.

28. Statistical Methods – adjustment for covariates

As described above, analyses will be conducted adjusting for baseline levels of outcomes (where available) and the stratifying variable of physiotherapist. For regression models that include data from both 3 and 9 months as the outcome, month and an interaction between month and randomised group will also be included.

In additional sensitivity analyses of the primary and secondary outcomes, any baseline characteristics that appear to exhibit a meaningful imbalance between treatment groups will be adjusted for, and the sensitivity of the conclusions drawn as a result of these additions will be assessed.

29. Statistical Methods – sensitivity analyses

The primary outcomes will be analysed within the non-inferiority framework. To assess the sensitivity of the conclusions of the assessment of non-inferiority, an inverse probability weighting approach will be used to

estimate the treatment effect for the two primary outcomes.[12]Inverse probability weights will be calculated for those participants who adhere to their assigned treatment, and models as described in Section 27 will be fit to the primary outcomes, using the weighted per-protocol dataset, which will include only those participants who attended at least 3 of their physiotherapy consultations.

30. Statistical Methods – subgroup analyses

Irrespective of the outcomes of the non-inferiority analysis, we will conduct exploratory analyses to evaluate moderation of the effect of video-conferencing versus face-to-face consultations on primary outcomes or on participant time and travel (secondary outcome) by pre-specified potential moderators, i) experience with online video platforms at baseline; ii) geographical residence; iii) beliefs about physiotherapy care delivery at baseline; and iv) confidence using technology at baseline. This will be assessed by including appropriate interaction terms between the moderators and the intervention term, where the superiority framework will be applied for interpreting results.

The *a priori* hypotheses to be tested are:

i) Participants who are less frequent users of video platforms at baseline will have less improvement in primary outcomes with video consultations (relative to face-to-face), compared to participants who are more frequent users.
ii) Participants who don't live in major city areas will report reduced participant-level time and travel distance (secondary outcomes) with video consultations (relative to face-to-face), compared to participants who live in major cities.
iii) Participants who believe that video consultations are less effective for managing musculoskeletal problems at baseline will report less improvement in primary outcomes with video consultations (relative to face-to-face), compared to those who believe video consultations are more effective.
iv) Participants who are less confident with using technology at baseline will report less improvement in primary outcomes with video consultations (relative to face-to-face), compared to participants who are more confident.

Models for these analyses will be as described in Section 27, including terms for the moderator and interactions between the moderator and each of randomised group and time, and the three-way interaction between moderator, time and randomised group. For hypotheses i, iii and iv, both primary outcomes will be analysed; for hypothesis ii, only participant time and distance travelled outcomes will be analysed. Results will be reported for the 3-month time point only.

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

Baseline characteristics of participants with one or both primary outcomes missing at 3 months will be compared to those of participants with both primary outcomes, as outlined in Section 25. If more than 5% of participants have at least one primary outcome missing at 3 months, multiple imputation will be applied. Missing baseline characteristics will be imputed using single mean imputation. Missing outcome values will be imputed separately by treatment group, using chained equations and predictive mean matching, using the five nearest neighbours. Imputation models will include baseline levels of outcomes and baseline characteristics that appear to be different between participants who provide complete follow up data and participants who do not. Initially imputation models for all outcomes will be chained 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 plots of proportions for binary outcomes.

To assess the potential impact of the violation of the missing-at-random assumption on conclusions for the primary outcomes, a pattern-mixture approach (as in White et al[22]) will be applied. We will explore the impact of the violation of the missing-at-random assumption if the assumption was violated in all groups, two groups, or in one group only.

32. Additional Analyses

A health economist will oversee assessment of incremental direct and indirect costs of video consultations compared to face-to-face. Primary evaluation will be between-group difference in knee-related health care costs and quality-adjusted life years (QALYs) using generalised linear models to adjust for baseline. Non-inferiority in QALYs will be demonstrated if the lower bound of the two-sided 95% CI for between-group difference of the AQoL-6D is above -0.08 (half a SD). If one type of consultation is superior but costs more, QALYs will be calculated using area under the curve over 9 months. The incremental cost per QALY as the ratio of difference in mean cost to difference in mean QALYs, and net benefits as the difference in QALYs times the social value of a QALY minus the difference in cost, will then be calculated. If non-inferiority of QALYs is demonstrated, then inferiority in terms of cost (and net benefits) will be assessed if the lower bound of the two-sided 95% CI for between-group difference exceeds zero. In each case, the degree of confidence in the monetary value of incremental net benefits will be calculated from the one-sided p-values for the between group difference in the money value of a QALYs less costs and presented as an acceptability curve across range of threshold money values of a QALY. In a secondary analysis, the cost of participant time will be included as an additional cost component.

33. Harms

Adverse events are reported and analysed as a secondary outcome- see Section 26.

34. Statistical Software

Stata v17 or later will be used (StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC)

35. References

1. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur. J. Pain.* 2004; 8(4):283-291.
2. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain.* 2001; 94(2):149-158.
3. Bellamy N, Carette S, Ford P, et al. Osteoarthritis antirheumatic drug trials.II. Tables for calculating sample size for clinical trials. *J. Rheumatol.* 1992; 19(3):1992.
4. Zhang H, Zhang K, Zhang X, et al. Comparison of two hyaluronic acid formulations for safety and efficacy (CHASE) study in knee osteoarthritis: a multicenter, randomized, double-blind, 26-week non-inferiority trial comparing Durolane to Artz. *Arthritis Res. Ther.* 2015; 17:51.
5. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J. Rheumatol.* 2000; 27(11):2635-2641.
6. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis.* 2005; 64(1):29-33.
7. Bennell KL, Nelligan R, Dobson F, et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain: A Randomized Trial. *Ann. Intern. Med.* 2017; 166(7):453-462.
8. Bennell KL, Campbell PK, Egerton T, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: A randomised clinical trial. *Arthritis Care Res (Hoboken).* 2016.
9. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA.* 2012; 308(24):2594-2604.

10. Wasserstein R, Schirm A, Lazar N. Moving to a World Beyond "p < 0.05". *The American Statistician.* 2019; 73(Supp 1):1-19.
11. Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? *Trials.* 2018; 19(1):499.
12. Mo Y, Lim C, Watson JA, White NJ, Cooper BS. Non-adherence in non-inferiority trials: pitfalls and recommendations. *BMJ.* 2020; 370:m2215.
13. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010; 340:c869.
14. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. *J. Rheumatol.* 1997; 24(4):768-778.
15. 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. *J. Clin. Epidemiol.* 2003; 56(2):138-147.
16. Martin K, Rejeski W, Miller M, James M, Ettinger Jr W, Messier S. Validation of the PASE in older adults with knee pain and physical disability. *Medicine & Science in Sports & Exercies.* 1999; 31(5):627-633.
17. 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.
18. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology.* 1989; 36(2):223-233.
19. Wiens BL. Multiple comparisons in non-inferiority trials: Reaction to recent regulatory guidance on multiple endpoints in clinical trials. *J Biopharm Stat.* 2018; 28(1):52-62.
20. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ.* 1998; 316(7139):1236-1238.
21. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology.* 1990; 1(1):43-46.
22. White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. *Stat. Med.* 2011; 30(27):3192-3207.

Appendix: Planned tables

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

	Group 1 (n=xxx)	Group 2 (n=xxx)
Recruitment source, n (%)		
Electronic (social media/email)		
Radio/newspaper		
Word of mouth/other		
Electronic data collection mode, n (%)		
Age (years)		
Female, n (%)		
Male, n (%)		
Height (m)		
Body mass (kg)		
Body mass index (kg/m^2)		
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		
Unilateral symptoms, n (%)		
Problems in other joints, n (%)		
Head		
Neck		
Back		
Hip		
Ankle		
Shoulder		
Elbow		
Hand/wrist		
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
Treatment expectations, n (%)[b]
No effect
Minimal improvement
Moderate improvement
Large improvement
Complete recovery
Confidence using technology in day to day life, n (%)[c]
Frequent user of technology, n (%)[d]
Mobile phone
SMS
Computer (desktop or laptop)
Email
Tablet
Online video platforms
Online shopping
Internet for any purpose
Social media
Beliefs that physiotherapy care delivery is effective, n (%)[e]
Face-to-face consultations with a physiotherapist
Group classes led by a physiotherapist
Telephone consultations with a physiotherapist
Video consultations with a physiotherapist
Home visits by a physiotherapist
Physiotherapy consultations in private physiotherapy clinics
Physiotherapy consultations in community health centres
Physiotherapy consultations in private hospital departments
Physiotherapy consultations in public hospital departments
Prior experience with physiotherapy care delivery, n (%)[f]
Face-to-face consultations with a physiotherapist
Group classes led by a physiotherapist
Telephone consultations with a physiotherapist
Video consultations with a physiotherapist
Home visits by a physiotherapist
Physiotherapy consultations in private physiotherapy clinics
Physiotherapy consultations in community health centres
Physiotherapy consultations in private hospital departments
Physiotherapy consultations in public hospital departments

[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 confidant, and extremely confident. Dichotomised into less confident (not at all and somewhat confident) and more confident (moderately and extremely confident).
[d]rated using 6-point Likert scales with response options of never, once every few months, once a month, once a week, several times a week, every day. Dichotomised into infrequent (never, once every few months, once a month) and frequent (once a week, several times a week, every day) users.
[e]rated using 4-point Likert scales with response options of not effective, somewhat effective, moderately effective and highly effective. Dichotomised into ineffective (not or somewhat effective) and effective (moderately and highly effective).
[f]rated as yes or no based on prior experience.

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

	Baseline		3 months		9 months	
	Group 1 (n=xxx)	**Group 2 (n=xxx)**	**Group 1 (n=xxx)**	**Group 2 (n=xx)**	**Group 1 (n=xx)[#]**	**Group 2 (n=xx)[‡]**
Primary outcomes						
Average pain on walking (NRS)						
Physical function (WOMAC)						
Secondary outcomes						
Health-related quality of life (AQoL-6D)						
Physical activity (PASE)						
Self-efficacy (ASES)						
Therapeutic alliance (WAI)[†]- patient rating						
Therapeutic alliance (WAI)[†]- therapist rating						
Consultation satisfaction[ß]						
Consultation convenience[ƒ]						
Consultations attended[*]						
Adherence to strengthening program[‡]						
Adherence to physical activity plan[‡]						
Strengthening exercise sessions completed[§]						
Participant time[¶]						
Physiotherapist time[µ]						
Participant distance travelled[£]						

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function); AQoL-6D=Assessment of Quality of Life instrument, (-0.04-1.0; higher scores indicate better quality of life); PASE=Physical Activity Scale for the Elderly (0->400; higher scores indicate better physical activity); ASES= Arthritis Self-Efficacy Scale (1-10; higher scores indicate greater self-efficacy); WAI= Working Alliance Inventory short form (12-84; higher scores indicate a stronger therapeutic alliance).

[†]Rated at 3 months only by patients & after the 5th consultation (or on the day the 5th consultation was due to be scheduled for participants who cancel/do not attend) by physiotherapists.

[ß]Rated using an 11-point numeric rating scale where 0=extremely unsatisfied to 10=extremely satisfied, at 3 and 9 months only.

[ƒ]Rated using an 11-point numeric rating scale where 0=extremely inconvenient to 10=extremely convenient, at 3 and 9 months only.

[*]As recorded by physiotherapists in treatment notes.

[‡]Rated by participants using an 11-point numeric rating scale where 0=strongly disagree to 10=strongly agree, at 3 and 9 months only.

[§]Reported by participants in whole numbers over the past week, at 3 and 9 months only.

[¶]Recorded by participants in log-book as total time spent per consultation (travelling to/from, waiting and consultation time).

[µ]Recorded by physiotherapists in treatment notes as total time spent per consultation (excluding note-taking and appointment scheduling).

[£]Recorded by participants in log-book.

Table 3: Change from baseline within groups on continuous secondary outcomes that were measured across time from baseline, and difference in change between groups (adjusted for baseline level of outcome, time, physiotherapist, and multiple measurements per participant).

	Mean (SD) change within groups				Difference in change between groups			
	Baseline minus month 3		Baseline minus month 9		Baseline to month 3		Baseline to month 9	
	Group 1 (n=xx)	Group 2 (n=xx)	Group 1 (n=xx)	Group 2 (n=xx)	Mean difference (95% CI)	P-value	Mean difference (95% CI)	P-value
Primary outcomes								
Average knee pain on walking (NRS)†								
Physical function (WOMAC)†								
Secondary outcomes								
Health-related quality of life (AQoL-6D)‡								
Physical activity (PASE)‡								
Self-efficacy (ASES)‡								

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function); AQoL-6D=Assessment of Quality of Life instrument, (-0.04-1.0; higher scores indicate better quality of life); PASE=Physical Activity Scale for the Elderly (0->400; higher scores indicate better physical activity); ASES= Arthritis Self-Efficacy Scale (1-10; higher scores indicate greater self-efficacy).

†For change within groups, positive changes indicate improvement. For difference in change between groups, positive differences favour Group 2.

‡For change within groups, negative changes indicate improvement. For difference in change between groups, negative differences favour Group 2.

Table 4. Between-group differences in secondary outcomes that were measured at follow-up time-points only.

Measures*	Mean (95% CI) difference at 3 months	P value	Mean (95% CI) difference at 9 months	P value
Therapeutic alliance (WAI)[†#]- patient rating			**NA**	**NA**
Therapeutic alliance (WAI)[†#]- therapist rating			**NA**	**NA**
Consultation convenience (0-10)[ƒ#]			**NA**	**NA**
Consultations attended[*#]			**NA**	**NA**
Consultation satisfaction[ß#]				
Adherence to strengthening program[‡#]				
Adherence to physical activity plan (0-10)[‡ #]				
Strengthening exercise sessions completed[§#]				
Participant time (min)[¶ ¥]			**NA**	**NA**
Physiotherapist time (min)[µ¥]			**NA**	**NA**
Participant distance travelled (km)[£¥]				

WAI= Working Alliance Inventory short form (12-84; higher scores indicate a stronger therapeutic alliance).
[†]Rated at 3 months by patients & after the 5th consultation (or on the day the 5th consultation was due to be scheduled for participants who cancel/do not attend) by physiotherapists.
[ƒ]Rated using an 11-point numeric rating scale where 0=extremely inconvenient to 10=extremely convenient, at 3 and 9 months.
[*]As recorded by physiotherapists in treatment notes.
[ß]Rated using an 11-point numeric rating scale where 0=extremely unsatisfied to 10=extremely satisfied, at 3 and 9 months only.
[‡]Rated by participants using an 11-point numeric rating scale where 0=strongly disagree to 10=strongly agree, at 3 and 9 months.
[§]Reported by participants in whole numbers over the past week, at 3 and 9 months.
[¶]Recorded by participants in log-book as total time spent per consultation (travelling to/from, waiting and consultation time).
[µ]Recorded by physiotherapists in treatment notes as total time spent per consultation (excluding note-taking and appointment scheduling).
[£]Recorded by participants in log-book.
[#]Positive between-group differences favour Group 2.
[¥]Negative between-group differences favour Group 2.

Table 5: Binary secondary outcomes and adjusted relative risks and risk differences. Counts and proportions based on complete case data.

	Month 3						Month 9					
	Group 1 (n=xx)	**Group 2 (n=xx)**	**Relative Risk (95% CI)**	**P-value**	**Risk Difference (95% CI)***	**P-value**	**Group 1 (n=xx)**	**Group 2 (n=xx)**	**Relative Risk (95% CI)**	**P-value**	**Risk Difference (95% CI)***	**P-value**
Improved pain[†]												
Improved function[†]												
Improved physical activity[‡]												
Purchased oral/topical medications[Ω]												
Consulted a health professional[Ω]												

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

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

[Ω]Participants who purchased oral/topical medication, or saw a health professional, at least once for their knee pain in the prior 3 months (excluding consultations delivered as part of intervention).

*Risk differences > 0 and relative risks > 1 favour Group 2

Table 6. Adverse events according to group, presented as number (%) of participants who had events.

	3 months		9 months	
	Group 1 (n=xx)	**Group 2 (n=xx)**	**Group 1 (n=xx)**	**Group 2 (n=xx)**
Adverse events:				
N reporting ANY adverse event				
Knee pain				
Pain in other areas				
Hip pain				
Ankle/foot pain				
Knee swelling				
Calf pain				
Knee stiffness				
Serious adverse events*				

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

Supplementary Table x: Medication use, doctor, and other health professional consultations according to group, presented as number (%) of participants who took medication or saw doctors or health professionals at least once, unless otherwise indicated.

	Baseline		3 months		9 months	
	Group 1 (n=xx)	Group 2 (n=x)	Group 1 (n=xx)	Group 2 (n=x)	Group 1 (n=xx)	Group 2 (n=x)
Medication use:						
N purchased ANY oral/topical medication						
Non-steroidal anti-inflammatories						
Analgesia (includes paracetamol combinations)						
Analgesia (opioids)						
COX-2 inhibitors						
N consulted ANY health professional						
Doctor consultations:						
General practitioner						
Rheumatologist						
Orthopaedic surgeon						
Sport & exercise physician						
Other medical specialist						
Other health professional consultations:						
Physiotherapist*						
Dietician						
Chiropractor						

Podiatrist	
Massage therapist	
Psychologist	
Osteopath	
Acupuncturist	
Other health professional	
Total number of health professional visits (mean (SD))	

*including both hospital and community settings but excluding consultations delivered as part of intervention; COX-2= cyclooxygenase-2.

Supplementary Table x. Baseline characteristics of participants who provided both primary outcomes at 3 months (completers) and those who did not provide both (non-completers).

	Non-completers (n=xxx)	Completers (n=xxx)	P-value
Group, n (%)			
Video consultations			
Face-to-face consultations			
Recruitment source, n (%)			
Electronic (social media/email)			
Radio/newspaper			
Word of mouth/other			
Electronic data collection mode, n (%)			
Age (years)			
Female, n (%)			
Male, n (%)			
Height (m)			
Body mass (kg)			
Body mass index (kg/m^2)			
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			
Unilateral symptoms, n (%)			
Problems in other joints, n (%)			
Head			
Neck			
Back			
Hip			
Ankle			
Shoulder			
Elbow			
Hand/wrist			
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	
Treatment expectations, n (%)[b]	
No effect	
Minimal improvement	
Moderate improvement	
Large improvement	
Complete recovery	
Confidence using technology in day to day life, n (%)[c]	
Belief that face-to-face physiotherapy care delivery is effective, n (%)[e]	
Belief that virtual physiotherapy care delivery is effective, n (%)[e]	
Prior experience with face-to-face physiotherapy care delivery, n (%)[f]	
Prior experience with virtual physiotherapy care delivery, n (%)[f]	
Average pain on walking (NRS)	
Physical function (WOMAC)	
Health-related quality of life (AQoL-6D)	
Physical activity (PASE)	
Self-efficacy (ASES)	

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function); AQoL-6D=Assessment of Quality of Life instrument, (-0.04-1.0; higher scores indicate better quality of life); PASE=Physical Activity Scale for the Elderly (0->400; higher scores indicate better physical activity); ASES= Arthritis Self-Efficacy Scale (1-10; higher scores indicate greater self-efficacy).

[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 confidant, and extremely confident. Dichotomised into less confident (not at all and somewhat confident) and more confident (moderately and extremely confident).

[d]rated using 6-point Likert scales with response options of never, once every few months, once a month, once a week, several times a week, every day. Dichotomised into infrequent (never, once every few months, once a month) and frequent (once a week, several times a week, every day) users.

[e]rated using 4-point Likert scales with response options of not effective, somewhat effective, moderately effective and highly effective. Dichotomised into ineffective (not or somewhat effective) and effective (moderately and highly effective).

[f]rated as yes or no based on prior experience.

Supplementary Table x: Adjusted per-protocol analysis of primary outcomes for assessment of non-inferiority hypotheses

	Mean (SD) change within groups Baseline minus 3 months		Difference in change between groups Group 2 minus Group 1	
	Group 1 (n=xx)	Group 2 (n=xx)	Mean difference (95% CI)	P-value
Primary outcomes				
Average knee pain on walking (NRS)†				
Physical function (WOMAC)†				

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function).
†For change within groups, positive changes indicate improvement. For difference in change between groups, positive differences favour Group 2.

Supplementary Table x: Moderation of the treatment effect on the primary outcomes at 3 months by frequency of use of video platforms, belief about video consultation effectiveness, and confidence in using technology; and moderation of the treatment effect on participant time and travel distance by location.

Outcome	Moderator	Mean (SD) change within Group 1	Mean (SD) change within Group 2	Group 2 minus Group 1 (95%CI)	Interaction P-value
Average knee pain on walking (NRS)	Frequency of use of video platforms				
	Less frequent				
	More frequent				
	Belief that video consultations are effective				
	No				
	Yes				
	Confidence in using technology				
	Not confident				
	Confident				
Physical function (WOMAC)	Frequency of use of video platforms				
	Less frequent				
	More frequent				
	Belief that video consultations are effective				
	No				
	Yes				
	Confidence in using technology				
	Not confident				
	Confident				
Participant time	Location				
	Major city area				
	Outside of major city area				

Participant travel distance	Location	
	Major city area	
	Outside of major city area	

NRS=numerical rating scale (0-10; higher scores indicate worse pain); WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale (0-68; higher scores indicate worse function).