Statistical Analysis Plan (SAP)

The EPIPHA-KNEE trial: Explain Pain to target unhelpful pain beliefs to Increase PHysical Activity in KNEE osteoarthritis – a multicentre, randomised controlled trial for people with painful knee osteoarthritis.

SECTION ONE: ADMINISTRATIVE INFORMATION

Title

The EPIPHA-KNEE trial: Effectiveness of an individualised pain science education and exercise program for knee osteoarthritis relative to best practice care – a randomised controlled trial.

Trial registration

Prospectively registered (Australian New Zealand Clinical Trial Registry Trial ID: ACTRN12620001041943; 2021)

SAP Version

Version: 2.0. Date: 17.09.2024

Protocol

This document has been written based on information contained the EPIPHAKNEE study protocol which was published as follows:

Stanton TR, Braithwaite FA, Butler D, Moseley GL, Hill C, Milte R, Ratcliffe J, Maher C, Tomkins-Lane C, Pulling BW, MacIntyre E, Esterman A, Stanford T, Lee H, Fraysse F, Metcalf B, Mouatt B, Bennell K. The EPIPHA-KNEE Trial: Explaining Pain to target unhelpful pain beliefs to Increase PHysical Activity in KNEE osteoarthritis – a protocol for a multicentre, randomised controlled trial with clinical- and cost-effectiveness analysis. *BMC Musculoskeletal Disorders* 2021; 22: 738. DOI:10.1186/s12891-021-04561-6

SAP Revisions

Not applicable

Names and affiliations

Document prepared by Associate Professor Tasha Stanton and Dr Ty Stanford, University of South Australia and Professor Kim Bennell, University of Melbourne

Emails: TR Stanton: tasha.stanton@unisa.edu.au; T Stanford: ty.stanford@unisa.edu.au; KL Bennell: k.bennell@unimelb.edu.au

Signatures:

Signature of senior statistician responsible (Dr Ty Stanford):  *Date:* 17.09.2024
Signature of chief investigator (A/Prof Tasha Stanton):  *Date:* 17.09.2024
Signature of chief investigator (Prof Kim Bennell):  *Date:* 17.09.2024

SECTION 2: INTRODUCTION

Background and Rationale

Symptomatic knee osteoarthritis is a leading cause of disability in older adults.[1] Structured physical activity, paired with education and weight management, is recommended as first-line care for knee osteoarthritis by all clinical practice guidelines [2] given their demonstrated benefits on pain and function.[3,4] Despite this, nine in ten people with painful knee osteoarthritis are physical inactive.[5] Interventions aimed at increasing physical activity levels, such as aerobic/strengthening exercise programs or behaviour change interventions (e.g., using coping skills) typically demonstrate short-term, but not long-term benefits.[6] Importantly, the decline in clinical benefit from short- to long-term timepoints of these programs is underpinned by declining physical activity level over time.[7] Thus, a key target to lasting clinical benefit is finding ways to sustain physical activity levels.

Recent work has shown that unhelpful, negative beliefs surrounding physical activity likely contribute to inactivity in people with knee osteoarthritis. People with osteoarthritis who are inactive are more likely to hold beliefs that they are physically unable to exercise, and that physical activity is unsafe relative to people with osteoarthritis who are active.[8,9] Additionally, people with knee osteoarthritis often hold beliefs that osteoarthritis is an incurable, progressive 'wear and tear' disease, with physical activity unlikely to be beneficial (or likely to be harmful) because their joint is 'bone-on-bone' (no cartilage left).[10,11] These beliefs are held despite evidence showing benefits of structured physical activity[3,4] even in those with severe, end-stage osteoarthritis, awaiting joint replacement.[12] In people with knee osteoarthritis, such unhelpful beliefs play a critically important negative role in their decision to engage with structured physical activity[10] and reduce their engagement in activities that may be pain-provoking.[13] Indeed, negative beliefs about osteoarthritis and activity predict lower future physical activity levels in people with knee osteoarthritis.[14]

Most current approaches that aim to increase physical activity level do not attempt to target these unhelpful beliefs, instead encouraging 'movement despite pain', with the promise of pain-relieving effects over time. Such approaches can arguably be counterintuitive to those with knee osteoarthritis if they consider pain to represent a marker of joint damage.[15] That is, if increases in pain during activity are construed as evidence that physical activity may not be safe for the joint, engagement in or adherence to physical activity programs is unlikely.[10] Further, if unhelpful beliefs are not challenged, then flare-ups of pain, stiffness, and/or swelling that commonly occur with initiation of a physical activity program in inactive individuals can and do derail long-term adherence.[16] While guidelines suggest inclusion of general education within physical activity programs, such education largely focuses on discussion of the health and joint-specific benefits of activity and how to perform activity accurately and safely.[17] Such education does not improve long-term activity adherence.[18]

Contemporary pain science education (PSE) aims to increase knowledge and target unhelpful and/or inaccurate beliefs surrounding pain *that underlie behaviour*, such as physical activity. PSE aims to reconceptualise what pain means to an individual.[15] Specifically, PSE aims to shift the conceptualisation of pain from that of a marker of tissue damage or injury to that of the need to protect the body from real or perceived threat.[15] By reconceptualising the

biological underpinnings of pain, PSE aims to enable an individual to move away from thinking that only structural pathology can contribute to pain. Instead, PSE explores the substantial central (and peripheral) neuroimmune adaptions that occur when pain persists and the impact of these changes on the relationship between pain and true tissue threat. In this way, PSE provides science-informed support for the fact that numerous factors can contribute to the experience of pain and as well as provide evidence that enhanced sensitivity can occur through central nervous system adaptions following pain persistence.[15] Such knowledge provides alterative hypotheses to pain or flare-ups experienced during activity: pain is not just about structural damage, but rather, can readily occur when the pain system is highly sensitised, as can occur in osteoarthritis.[19] High-level evidence from randomised controlled trials (RCTs) in numerous chronic pain conditions show that PSE increases pain knowledge and reduces unhelpful pain beliefs,[20,21] and clinical effects include reduced pain, improved function and disability, and increased physical activity levels in numerous chronic pain conditions, particularly when paired with an exercise program.[20-26]

Our team have adapted the PSE intervention to include osteoarthritis-specific pain science information, including the many central nervous system and immune system changes that occur with osteoarthritis[19,27,28] and the positive influence of physical activity on these system-level changes and on pain.[29-31] This was created into a patient-facing educational resource.[32] Feasibility work undertaken by our group shows positive effects of a physiotherapist-delivered PSE individualised education and walking program in people with painful knee osteoarthritis.[33] Specifically, the PSE and walking intervention demonstrated high levels of participant-rated credibility, feasibility, and acceptability, with promising effects on pain knowledge, pain intensity, function, and physical activity levels (as measured via wrist-worn accelerometry) in people with knee osteoarthritis.[33] Thus, transition to a full clinical trial to comprehensively test clinical effectiveness is warranted. As part of this transition, inclusion of a lower limb strengthening component to the walking program was deemed relevant to be consistent with best practice care.[2]

The aim of the EPIPHAKNEE Trial is to evaluate the clinical effectiveness of individualised pain science-informed walking, strengthening and education program (*PSE-Informed Care)* for improving severity of overall knee symptoms (pain, stiffness, function; via the WOMAC Osteoarthritis Index Total Score) and increasing physical activity levels (average step count) at 52 weeks post-randomisation compared to *Best Practice Care* (general education, and an individualised walking and strengthening program) in people with painful knee osteoarthritis. The present statistical analysis plan details the analysis for primary, secondary, and exploratory outcomes.

Objectives

Study Research Question:

Our primary research question is:

For people with clinically diagnosed knee osteoarthritis, is the *PSE-Informed Care intervention* more effective (greater improvement in severity of overall knee symptoms of

pain, stiffness and function; greater increase in physical activity levels via average step count) than the *Best Practice Care intervention* (control) at 52 weeks?

Research Hypotheses:

Our primary hypothesis is that the *PSE-Informed Care intervention* will lead to greater improvement in severity of overall knee symptoms (pain, stiffness, and function; via the WOMAC Osteoarthritis Index Total Score) and greater increases in physical activity levels (via average step count derived from wrist-worn accelerometry) than the *Best Practice Care intervention* (control) at 52 weeks.

Secondary hypotheses:

The *PSE-Informed Care intervention* will have greater benefits for severity of overall knee symptoms (pain, stiffness, and function; via the WOMAC Osteoarthritis Index Total Score) and physical activity levels (step count derived from wrist-worn accelerometry) than the *Best Practice Care intervention* at 12 and 26 weeks.

The *PSE-Informed Care intervention* will have greater benefits for other clinical outcomes (other pain intensity measures, physical function, average minutes of sedentary, light, moderate/vigorous activity using Troiano et al cut-points,[34,35] depression, anxiety, stress, self-efficacy, fear of movement, quality of life, pain catastrophising, knee perception, osteoarthritis/activity conceptualisation, global rating of change, proportion meeting minimal clinically important difference for primary outcomes) than the *Best Practice Care intervention* at 12, 26, and 52 weeks.

Exploratory hypotheses:

The *PSE-informed care intervention* will have greater benefits for other physical activity outcome measures derived from the wrist-worn accelerometry (intensity outcomes using Smuck et al cut-points,[36] volume outcomes using Troiano et al cut-points[34,35] and Smuck et al cut-points,[36] 24 hour time-use composition expressed as isometric log ratios,[37,38] and sleep duration using van Hees et al algorithm[39]) than the *Best Practice Care intervention* at 12, 26, and 52 weeks.

SECTION 3: TRIAL METHODS

Trial design

The EPIPHAKNEE Trial is a two-group, superiority parallel-design randomised controlled trial. Participants and outcome assessors are blinded to study group and hypotheses. Ethical approval was attained from the University of South Australia Human Research Ethics Board (ID202377), The Central Adelaide Local Health Network (CALHN) Human Research Ethics Board (ID12579), The University of Melbourne Human Research Ethics Board (ID2057540), and Flinders University Human Research Ethics Committee (ID4478).

Randomisation

Randomisation used a 1:1 allocation ratio. On completion of the baseline assessment, participants were randomly allowed into one of the two groups: PSE-Informed Care (PSE + individualised exercise program); ii) Best Practice Care (standard education + individualised exercise program). A biostatistician with no involvement in the conduct of the trial created a computer-generated randomisation schedule Research Randomizer (https://www.randomizer.org/). Within sex and treatment site (Adelaide, Melbourne) strata, allocation was applied using permuted blocks of varying size (4 and 6, randomly). At time of enrolment, the trial coordinator who was not involved in outcome assessment or treatment provision used REDCap to randomly allocate participants to their treatment group (based on the sex and treatment site specific strata randomisation schedule). The trial coordinator then set-up treatment appointment with the appropriate physiotherapist (given participant location preference and group assignment).

Sample size

A total of 198 participants (n=99/group) was required to have at least 90% power to detect an effect size (d) of 0.5 (the effect size recommended by the Osteoarthritis Research Society International as the minimally clinically important difference for outcomes in osteoarthritis trials) in severity of osteoarthritis symptoms (Western Ontario and McMaster Universities Arthritis Index; WOMAC total score) and physical activity levels (step count calculated by wrist-worn accelerometry) at 52-weeks post randomisation. The sample size was calculated for an interaction effect between time (4 observations) and levels of intervention (2 groups), using an estimated inter-observation correlation of 0.7, and adjusted for up to 20% loss to follow-up and clustering based on therapist (~12 therapists/group, ~6.5 participants/therapist; ICC = 0.05).

Framework

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

Statistical Interim analyses and stopping guidance

Nil.

Timing of final analysis

Final analyses will be performed after all participants have reached the 12-month (primary) timepoint and completed their outcome measures.

Timing of outcome assessments

Primary and secondary outcomes are collected at baseline (0 weeks), 12 weeks, 26 weeks, and 52 weeks. Treatment expectancy, treatment credibility, and intention to exercise are assessed after participants receive the first treatment session.

SECTION 4: STATISTICAL PRINCIPLES

Level of statistical significance

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

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 symptoms (WOMAC total score) and physical activity levels (via wrist-worn accelerometry). We will not adjust for multiple outcomes but will instead report all treatment estimates, confidence intervals, and p-values in order to let readers make their own judgement about the relative weight of the conclusions on the effect of the intervention (i.e., at the level of Type I error control should they wish, at risk of making Type II errors).[40,41] This approach aligns with the use of p-values favoured by the American Statistical Association.[42] Additionally, s-values (surprisal or Shannon information) will be, at a minimum, provided in supplementary material to popularise critical thinking about p-values and provide alternative contextual information about the weight of evidence against the null hypotheses.[42,43]

Confidence intervals to be reported

All confidence intervals will be 95% confidence intervals, consistent with the (two-sided) significance level of 0.05.

Adherence and Protocol Deviations

The primary analysis will be based on the principle of intention-to-treat, whereby participants are included in the groups to which they were originally assigned, regardless of their adherence to their assigned treatments. Any protocol deviations (if they occur), including errors applying inclusion/exclusion criteria and/or administration of the wrong intervention will be summarised in trial results (patient flow diagram/text) by treatment group. Randomisation errors resulting from these errors will be handled according with contemporary recommendations.[44]

Multiple measures of adherence to the two active interventions are used in this trial and data from all measures will be reported using means, standard deviations (or medians and quartiles) or number counts and percentage as appropriate for each intervention group. Measures of adherence include:

- Attendance at in-person and telehealth treatment sessions (where attendance at 8 or more sessions out of 11 sessions will be considered acceptable adherence).
- Self-rated adherence to the walking program (11-point NRS; strongly agree to strongly disagree) at each follow-up timepoint (12, 26, and 52 weeks) for the recommended: number of walks per week; distance/time of each walk; and intensity of each walk (where 8 or higher ratings on the 11-point scale will be considered acceptable self-reported adherence to the walking program).

- Self-rated adherence to the strengthening program 11-point NRS; strongly agree to strongly disagree) at each follow-up timepoint (12, 26, and 52 weeks) for the recommended: number of days per week and sets/reps of each exercise (where 8 or higher ratings on the 11-point scale will be considered acceptable self-reported adherence to the strengthening program).
- Self-rated number of days over the past week they engage in the walking program and in the strengthening program at each follow-up timepoint (12, 26, and 52 weeks). Because each person will differ in the number of days they are meant to be engaging in the exercise programs each week (and at each timepoint), a set acceptable value will not be used, rather the raw data will be reported.

COVID-19 impact on adherence to physical activity

Given data collection of this study was situated within COVID-19 early pandemic and within Australia (which underwent strict activity restrictions), participants rated the degree to which COVID-19 and related activity restrictions influenced their usual physical activity and/or exercise routines over the past 4 weeks. This was assessed at baseline, 12, 26, and 52 weeks using a 7-point Likert rating scale ranging from a large reduction in usual physical activity levels/exercise routines (1) to a large increase in usual physical activity levels/exercise routines (7). We will present the frequency distributions of each intervention group for these outcomes as well as the proportion of participants in each group reporting sizable reductions in usual physical activity levels due to COVID-19 (e.g., ≤ 2, representing moderate to large reductions in activity).

SECTION 5: TRIAL SAMPLE

Screening data

Screening data will be collected and summarised. A CONSORT flow diagram will be used.[45] 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.

Eligibility

Trial inclusion and exclusion criteria are described in the trial protocol. Reasons for exclusion will be summarised in the CONSORT[45] flow diagram.

Recruitment

A CONSORT flow diagram[45] will be used to describe the number of people enrolled, randomised, allocated to each treatment group, lost to follow-up (including reasons), and analysed.

Withdrawal/follow-up

If a participant withdraws from the study, the nature, timing of and reasons for withdrawal will be described (provided the participants responds to the research team). Any data

provided up to the point of withdrawal will be analysed in accordance with intention to treat analyses, unless the participant specifically requests to withdraw their data from the study. Losses to follow-up (including reasons) will be summarised in the CONSORT flow diagram by treatment group.

Baseline characteristics

Baseline characteristics will be summarised by treatment group and presented in a table:

- Age (years)
- Sex (female, male, other)
- Height, body mass, body mass index (m, kg, m/kg^2, respectively)
- Level of education (Did not complete high school, completed high school, non-university qualification, university qualification [Bachelors], University post-graduate degree)
- Socioeconomic status; classified based on residential postcode using Australian Bureau of Statistics Index of Relative Socioeconomic Advantage and Disadvantage [IRSAD] score for Socio-economic Indices for Areas [SEIFA], ranging from 1 (most disadvantaged) to 5 (most advantaged).
- Duration of knee OA symptoms (months, years)
- Unilateral symptoms (yes, no)
- Problems in other joints (select: hand, shoulder, neck, back, hip, ankle/foot; total painful sites [count])
- Current medication use (select: acetaminophen, oral and topical anti-inflammatory drugs, oral corticosteroids, oral opioids; any oral pain medication [count])
- Interventions received in the past month (interventions provided as a list [count])
- Co-morbidities (via Functional Co-Morbidity Index[46])

Baseline characteristics will be summarised using mean and standard deviation for continuous and empirically symmetrically distributed variables (median and inter-quartile range will be used for empirically asymmetric distributed continuous variables), and count and percentage of observations for categorical outcomes. Test of statistical significance will not be undertaken for comparing baseline characteristics of treatments groups; rather the clinical importance of any imbalance will be noted.[45]

An appendix table will provide summaries of baseline characteristics and baseline levels of primary and secondary outcomes and compare these characteristics between those participants who provide both primary outcomes at 52 weeks, and those participants who are missing one or both primary outcomes at 52 weeks. *t*-tests will be used to compare the means for continuous characteristics/variables between the level of missingness groups (or permutation tests in the case of obvious empirical asymmetry), and Fisher's Exact tests will be used to compare the distributions of categorical characteristics between the level of missingness groups.

SECTION 6: OUTCOME VARIABLES

Outcome definitions.

Primary outcomes:

- Severity of overall knee symptoms: Average knee pain, stiffness, and function over the last 48 hours is assessed using a self-rated total Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index score.[47] The total WOMAC score is a continuous measure that ranges from 0 (no symptoms or dysfunction) to 96 (maximal symptoms and dysfunction).
- Physical activity level (step count): Average daily step count over the last 7 consecutive days is assessed using wrist-worn accelerometry (Actigraph GT9X Link, Pensacola, FL). Data are collected from the device in 60 s epochs. Participants complete a paper-based activity log to capture sleep and wake times, as well as times the accelerometer was removed and then placed back on. Valid wear time is operationalised as having ≥10 h of waking wear time for a minimum of 4 days (3 weekdays and 1 weekend day).[34,35] Average daily step count is calculated as the weighted average (5:2) of weekday and weekend step counts.

Secondary outcomes

- Knee pain intensity – overall (numerical rating scale): Average knee pain over the past week is self-assessed using an 11-point numerical rating scale (NRS) with anchors of 'no pain at all' (score 0) and 'worst pain possible' (score 10).[48]
- Knee pain intensity – walking (NRS): Average knee pain over the past week during walking is self-assessed using an 11-point numerical rating scale (NRS) with anchors of 'no pain at all' (score 0) and 'worst pain possible' (score 10).[48]
- Knee pain intensity (WOMAC): Average knee pain over the past 48 hours is self-assessed using the WOMAC pain subscale.[47]
- Knee function (WOMAC): Average knee function over the past 48 hours is self-assessed using the WOMAC function subscale.[47]
- Physical activity (average daily minutes using Hildebrand cut-points[49]): Average daily minutes of sedentary, light, moderate, and vigorous activity from the wrist-worn accelerometry will be used.
- Depressive symptoms: The 4-item PROMIS depression subscale will be used to assess severity of depressive symptoms over the past week.[50]
- Anxiety: The 4-item PROMIS anxiety subscale will be used to assess severity of anxiety over the past week.[50]
- Stress: The 4-item Perceived Stress Scale will be used to assess stress over the past month.[51,52]
- Fear of movement: The 6-item Brief Fear of Movement Questionnaire will be used.[53]
- Pain catastrophising: The brief 4-item Pain Catastrophising Scale will be used.[54]
- Pain self-efficacy: The Pain Self-Efficacy Questionnaires (PSEQ) will be used.[55]

- Conceptualisation of osteoarthritis/activity: The osteoarthritis and activity conceptualisation scale will be used.[56]
- Health-related quality of life: The EQ-5D-5L[57] will be used.
- Knee perception: The Fremantle Knee Awareness Questionnaire (FreKAQ)[58] will be used.
- Global rating of change overall in study knee at 12, 26 and 52 weeks: This will be scored using a 7-point ordinal scale[59] with response options of "Much Worse" to "Much Better". Responses will be dichotomised to "Improved" ("Moderately Better" and "Much Better") and "Not improved" (all other choices). The proportion of participants in each group who are "improved" and "not improved" will be provided.
- Global rating of change overall in physical function at 12, 26 and 52 weeks: This will be scored using a 7-point ordinal scale[59] with response options of "Much Worse" to "Much Better". Responses will be dichotomised to "Improved" ("Moderately Better" and "Much Better") and "Not improved" (all other choices). The proportion of participants in each group who are "improved" and "not improved" will be provided.
- Global rating of change overall in physical activity level at 12, 26 and 52 weeks: This will be scored using a 7-point ordinal scale[59] with response options of "Much Worse" to "Much Better". Responses will be dichotomised to "Improved" ("Moderately Better" and "Much Better") and "Not improved" (all other choices). The proportion of participants in each group who are "improved" and "not improved" will be provided.
- Proportion who meet or exceed the minimal clinically important difference (MCID) for primary outcomes: the proportion of participants in each group whose improvements meet or exceed the MCID for primary outcomes of overall knee symptoms (a change of 11.5 points on the 0-96 WOMAC Total score[60]) and physical activity level (a change of ≥1000 steps; the smallest increase that has been shown to reduce the risk of functional decline in people with knee osteoarthritis[61]).

Exploratory outcomes:

- Physical activity (average daily minutes using Smuck et al. (2017)[36] cut-points): Average daily minutes of sedentary, light (separated into 3 categories: light 1, light 2, light 3), moderate, and vigorous activity from the wrist-worn accelerometry will be used.
- Physical activity (overall volume using Hildebrand cut-points[49]): Volume (intensity x time) from the wrist-worn accelerometry will be used.
- Physical activity (overall volume using Smuck et al. (2017)[36] cut-points): Volume (intensity x time) from the wrist-worn accelerometry will be used.
- 24 hour time-use compositional data (expressed as isometric log ratios)[37,38] from the wrist-worn accelerometry will be used.
- Sleep: average daily sleep time (minutes), as identified by the van Hees et al. algorithm[39] (using the self-reported sleep logs to guide the algorithm) from the wrist-worn accelerometry will be used.

Outcome reporting.

Table of raw means and standard deviations (or medians and IQRs/bootstrapped 95% confidence intervals; as appropriate) will be presented for continuous primary and secondary outcomes at baseline, 12 weeks, 26 weeks, and 52 weeks. Proportions (count, frequency) will be presented for the outcomes of Global Rating of Change and proportion meeting MCID for primary outcomes at 12 weeks, 26 weeks, and 52 weeks. These will be provided to characterise the unadjusted values observed for transparency but will likely be reported in supplementary materials as the model estimates (adjusted for covariates and time effects) are of primary interest. Note that in the event the specified statistic (median or count) is generated based on 3 or less observations, the statistic will be withheld for patient confidentially/identifiability.

Similarly model estimates (fixed effect coefficients and random effect variance/SDt estimates, model fit statistics and model diagnostics) will be publicly available (Github, or OSF.org) from analysis transcripts (see in subsequent sections about dynamically created verbatim markdown on analysis inputs and outputs or so called "literate programming").

As the analysis is aimed at providing clinical knowledge, the primary and secondary (modelled) outcomes will be presented as so called marginal means which provide estimates of expected outcome values (or contrasts) on the scale of the outcome variable (for continuous variables using the identity link function, or the probability of a binary outcome in the case of dichotomous outcomes when the inverse link function is applied to the relevant linear combination of model estimates).[62] Marginal means will be operationalised as model predictions using the average sample values of continuous model predictors further averaged over the sample observed joint distribution of categorical predictors (weighted by frequency). In the case of binary outcomes, averaging will occur on the link scale and then inverse transformed to the estimated probably.

The specific marginal means for primary reporting will time-point specific means (probabilities): i) per intervention group; ii) for differences from baseline within intervention groups; and, of primary interest iii) difference between intervention groups from baseline (differences of differences). All marginal means will be presented with 95% bootstrapped confidence intervals (case-wise, non-parametric quantiles) for robustness (free of distributional assumptions) and to quantify asymmetry in confidence bounds especially in the case of transformed outcomes.

Exploratory outcomes will be reported in a manner consistent with primary and secondary outcomes. These may be included within the published manuscript describing primary and secondary outcomes (e.g., as supplementary files), or they may be published as separate manuscripts, should additional analyses be undertaken (e.g., effect of reallocation of time[37]).

SECTION 7: ANALYSIS

Analysis methods – General considerations

The statistical analysis plan will be published on the Centre for Health, Exercise and Sports Medicine's website. Analyses will be performed using R and Stata by the biostatisticians (TS,

ES) blinded to treatment group details. All analysis decisions and reporting will be made prior to group unblinding.

Treatment effect estimates (marginal means as specified previously) and associated 95% bootstrapped confidence intervals will be provided, and all statistical tests will be 2-sided using an alpha level of 0.05. Model estimated changes from baseline will be presented for each group at each timepoint using the mean change and 95% confidence intervals.

If the proportion of missing data in the primary outcomes is equal to or exceeds 5%[63] at the timepoint of 52 weeks, missing outcome data (at all timepoints) will be imputed using two-level hierarchical multiple imputation methodology[64,65] and the subsequent analysis of these data will take the place of the primary analyses (observed data only analysis will become the sensitivity analysis), providing assumptions are reasonable.[66] Otherwise the observed/complete case data will be analysed as the primary analysis and the two-level hierarchical multiply-imputed data (if missingness exists) analysis will serve as the sensitivity analysis.

Analysis methods – Primary outcomes

We will analyse the effect of the treatment arms for the two outcomes independently. Linear mixed-effects modelling will be used to model participant outcome trajectories: overall knee symptoms (WOMAC total score) and physical activity level (average daily step count over 7 days).

All models will include adjustment for randomisation strata variables of sex (male/female) and study site (Adelaide/Melbourne) as fixed effects.

The intervention group will be modelled as a binary variable and its interaction with time will serve as an estimate of a mean differential effect between the intervention groups over time.

Random intercepts for participant and physiotherapist will be included and retained if the variance estimates are non-zero (to a practical level of estimate precision). The time variable (as a fixed effect) will initially be attempted to be operationalised as a continuous variable so that random slopes (correlated with their respective random intercepts) within participant can be fit and tested for significance via log-likelihood ratio tests (LRTs). If the slope random effect is significant, this provides evidence of heteroscedastic trajectories of the outcome within participant that is of practical interest to quantify the relationship between baseline values and follow-up values at the individual level. For example, are participants with lower starting physical activity values on average more likely to have shallower improvements (if the correlation between the random intercept and slope are positively correlated)?

Model diagnostics (importantly, linearity between conditional residuals vs predicted values – by time-point and group) will inform the appropriate operationalisation of the time variable. Figure 1 outlines the simple transformations of (continuous) time that will be considered, in

order left to right, with the final time variable considered reverting to a flexible 4-level categorical variable. If residual linearity is achieved using an untransformed or transformed continuous time variable but there exists apparent residual heteroscedasticity over time, two simple transformations of the strictly positive outcome variable(s) will be considered to create variance stabilisation. In the case variance stabilisation cannot be achieved using a (un)transformed continuous time variable, or residual linearity cannot be achieved using a (un)transformed continuous time variable; a flexible categorical follow-up time variable will be used. The previously mentioned random slopes will take the analogous form of (unstructured) correlated random intercepts within participant per time-point which can also be tested against an intercept only random effect model for parsimony via an LRT of competing models (and will also allow an assessment of random intercept homoscedasticity over time).

The entire model fitting, selection and diagnostic processes will be made available in supplementary files using literate programming principles[67,68] for reproducibility and transparency.

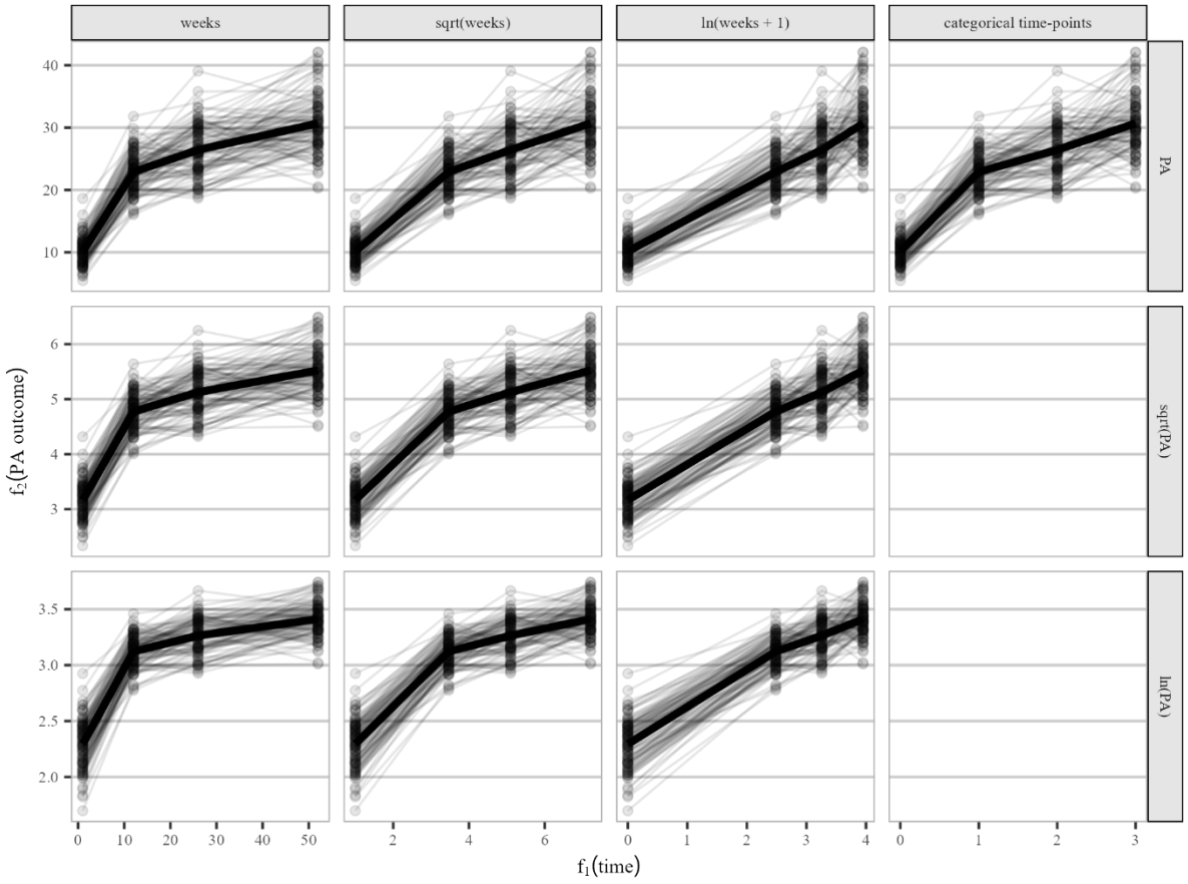

Figure 1: Conceptualisation of possible follow-up time transformations to linearise its relationship with the outcomes (if required) in addition to possible outcome transformations to create variance stabilisation in the case of residual variance being a function of time – made more likely as the outcome variables can only take non-negative values. Transformations are limited to simple monotonic functions for understandability and ease of reporting.

The primary analysis will report the adjusted mean change in outcome from baseline to 52-weeks post-randomisation between the PSE-informed care and control group (providing marginal means and 95% confidence interval). For consistency across marginal mean estimation and to provide appropriate nominal interval coverage in confidence intervals in the presence of (non-linear) transforms of the outcome variables (either directly applied to continuous outcomes or implicitly using the log-odds link for binary outcomes), the back-transformed marginal mean point estimates (on original scale) will be provided with non-parametric quantile, case-wise bootstrapped 95% confidence intervals. We will present confidence intervals (and unadjusted p-values) for all prespecified primary marginal means as well as the p-value associated with the interaction terms between time and intervention (*F*-test).

Please refer to subSAP_PA_v10.pdf for more technical details and guidance on the primary analysis modelling.

Analysis methods - Secondary outcomes

Linear mixed-effects models will also be used to estimate the effect of allocation to treatment group on all continuous secondary outcomes (using the same methodology as described above): knee pain intensity over the past week and past 48 hours, knee pain intensity walking over the past week, function, daily minutes of sedentary, light, moderate, and vigorous activity, depression, anxiety, stress, fear of movement, pain catastrophising, pain self-efficacy, OA/activity conceptualisation, health related quality of life, and knee perception.

For the binary secondary outcomes, participants experiencing improvements based on global change scores and those achieving the MCID in improvement in overall knee symptoms and physical activity, logistic mixed effects models (generalised linear mixed models with a binary outcome and using the log-odds link function) will be used to estimate the to the equivalent marginal mean probabilities based on (time, treatment group) combinations and contrasts as specified in Exemplar Table 2. Modelling will otherwise be equivalent with the modelling methodology (and covariates) as described for the primary analysis with the following caveats:

1. Marginal (mean) probabilities for (time, treatment group) combinations will be computed and averaged on the log-odds link/linear predictor scale before back transformation.
2. Marginal (mean) probability differences (time points within intervention group and differences between intervention groups based using the difference in time points within intervention group) will be computed as above separately for each for (time, treatment group) combination, then the point estimate will be the difference on the response (probability scale).
3. Random effects will be limited to random intercepts per person and physiotherapist and not include heterogeneity over time or random slopes because of the additional complexity in using non-identity link functions in generalised linear mixed effect models (and subsequent complexity in their interpretation).

Because of the number of secondary analyses and multiple comparison concerns, model/variable selection (if required) will be guided by goodness of fit statistics (BIC) (as opposed to variable significance) and confidence intervals of model estimates will be strongly preferred (over p-values, with the exception of the group by time overall interaction effect as previously specified).

Other exploratory variables that are not part of the main trial, but that were included in the original published protocol (tactile acuity, implicit motor imagery ability, pain conceptualisation drawing, and therapist related factors), will be used for other analyses and reported in separate publications.

Statistical Methods – adjustment for covariates

As described above, analyses will be conducted adjusting for covariates, which will include the two randomisation stratifying variables (sex, recruitment site).

Statistical Methods – sensitivity analyses

A sensitivity analysis will estimate treatment effects on the primary outcomes at 52 weeks assuming full adherence to the assigned treatment (where acceptable adherence is defined as attendance at 8 or more of 11 physiotherapy consultations). The same marginal mean estimates (as in the primary analysis) will be created for comparison with the primary analysis which differs by removing outcome values for participants who attended 7 or less consultations. Instrumental variable estimation will be used to estimate the complier-average causal effect (CACE), and either two-stage least squares[69-71] or Bayesian[72] models will be fit (as appropriate). Mean CACE of the intervention on severity of overall knee symptoms (WOMAC Total score) and on physical activity (average daily step count) will be reported with 95% confidence intervals and p-values. Propensity score weighting will be used to estimate the average treatment effect in the treated (ATET),[73,74] with mean, 95% confidence intervals and p-values reported.

A sensitivity analysis will be undertaken to consider the effect of COVID-19 activity restrictions on the primary outcomes. To address potential COVID-19 related effects, a three-way interaction between intervention group, time and time-varying indicator for COVID-19 restricted physical activity (based on self-reported COVID-19 activity impact, as detailed above) with group by time effect included as covariates for participants. Statistical significance (F-test for the null hypothesis of the additional coefficients being equal to 0) will be used to determine whether intervention COVID-19 related effects are statistically warranted (as either group by time-specific interaction effect or time-specific effect) and will be subsequently removed for model parsimony in the event of non-significance.

A sensitivity analysis will be undertaken controlling for any baseline variables that are unbalanced between groups, should the unbalanced baseline variables be of known relevance to prognosis or treatment response.

If the multiple imputation 5% missing at 52-weeks threshold is reached, multiply-imputed analyses will be the primary analysis and the complete-case analyses becomes the sensitivity analyses.

Statistical methods – subgroup analyses

We will conduct exploratory analyses to investigate whether the relative effects of each intervention on change in each of the primary outcomes at 52 weeks is moderated by: i) pain self-efficacy; ii) body perception; iii) self-regulated learning ability. Linear mixed models will include terms for treatment, the moderator variables, and an interaction between the two. For the continuous moderators below, tertiles/quartiles will be used for analysis; however, if domain accepted thresholds exist, these will be used and references provided. The estimated effects of treatment and 95% confidence intervals will be presented for each moderator, with visual representations included for continuous moderators.

The hypothesis, and rationale, for each chosen moderator is below:

- Pain self-efficacy: Participants in the intervention group who have higher self-efficacy at baseline (higher PSEQ scores[55]) will report greater improvement in primary outcomes than those who have lower self-efficacy (relative to control group). Past work has used severity bands where scores <20 represent severely impaired pain self-efficacy, 20-30 moderate impairment, 31-40 mild impairment, and >40 minimal impairment.[75] Dependent upon data spread, a dichotomised outcome may be used where ≥31 reflects low/mild severity and ≤30 reflects moderate/high severity.[75]
- Body perception (questionnaire): Participants in the intervention group who have less impaired body perception at baseline (lower FreKAQ scores[58]) will report greater improvement in primary outcomes than those who have more impaired body perception (relative to control group). Past work has shown that scores ≤17 predict response (pain relief) to exercise-based intervention.[76]
- Self-regulated learning ability (via 7 items from the Motivated Strategies for Learning Questionnaire; 7-point Likert scale):[77] Participants in the intervention group who have greater self-regulated learning ability at baseline (higher scores on the 7-items[77]) will report greater improvement in primary outcomes than those who have lower self-regulated learning ability (relative to control group).

Statistical methods – Trial process outcomes

Participant expectations of treatment outcomes are assessed after receiving the first intervention using a 5-point scale in response to the question "What effect do you think this treatment will have on your knee problem", with choices ranging from "No effect at all" to "Complete recovery". Treatment credibility is assessed after receiving the first intervention using the modified Credibility and Expectancy Questionnaire (CEQ).[78] Intention to exercise is assessed using the Intention to Exercise Scale[79] after the first intervention and at each outcome timepoint (12, 26, and 52 weeks). The mean group scores on the treatment expectancy scale, the modified CEQ, and the intention to exercise scale will be compared

between intervention groups using sample t-tests (or permutation tests in the case of apparent asymmetry).

Therapist perceived connection with the participant will be self-rated using a 100mm visual analogue scale of Connectedness (0 = no connection at all; 100 = maximum connection). The mean scores on the Connectedness Scale will be compared between intervention groups using independent sample t-tests (or permutation tests in the case of apparent asymmetry).

Blinding status of the participants and treating therapists will be undertaken via allocation guesses (and reasons), then forced choice.[80] Notably, all therapists providing the PSE-Informed Care intervention were unavoidably aware they were providing the intervention of interest in the trial (i.e., a new intervention requiring in-depth knowledge upskilling). Thus, we will aim to evaluate whether therapists providing the Control intervention similarly felt that they were providing the intervention of interest of the trial (i.e., similar therapist treatment expectancy effects). For trial participants, we will calculate the proportion of people who were successfully blinded (incorrect guess or unsure) for each intervention group. We will compare between intervention groups using Fisher's exact tests. We will also calculate the Bang Blinding Index (percentage of blinding beyond chance) for each intervention group,[80,81] whereby unblinding will be considered present if the 95% confidence interval does not include 0,[80] and unblinding beyond chance if the Blinding Index is >0.20.[82] For trial therapists, if ≥80% of therapists in the control group were unsure or incorrectly thought they provided the intervention of interest of the trial, we will consider therapist treatment expectancy effects to be adequately similar between groups.

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

Baseline characteristics of participants with one or both primary outcomes missing at 52 weeks will be compared to those of participants with both primary outcomes as described previously. If more than 5% of participants have at least one primary outcome missing at 52 weeks, multiple imputation will be applied for the primary analysis. The number of imputed datasets will be the same for all imputation analyses and equal to the number of participants with at least one missing primary outcome.[83] Missing baseline characteristics will be imputed using (single level) chained equations imputation. Missing outcomes values will be imputed using two-level imputation method as described previously. Initially imputation models for all outcomes (primary and secondary) 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 (MAR) assumption on conclusions for the primary outcomes, a pattern-mixture approach[84,85] will be applied. We will explore the impact of the violation of the missing-at-random assumption if the assumption was violated in both groups or in one group only.

Harms:

Data on adverse effects are actively captured via pre-defined questions that the treating physiotherapist completes after each intervention session (via Case Notes and REDCap attendance surveys). Data on adverse effects are also taken via passive capture from participants (self-report) at all assessment time points (0-52 weeks). Adverse effects will be reported using the United States Food and Drug Administration (FDA) definitions.[86] Specifically, an adverse effect is defined as 'any untoward medical occurrence associated with the intervention, whether or not considered related to the intervention.' A serious adverse effect is considered present when death, threat to life, in-patient hospitalisation or prolongation of existing hospitalisation, or persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions occurs (or medical intervention is required to prevent occurrence). Lists of all adverse effects and serious adverse effects reported throughout the trial period (0-52 weeks; available data for each participant) will be displayed and the proportion of participants in either group that experienced them will be calculated. We will specifically calculate the proportion of participants that experienced any related (or possibly related) adverse effect and any related (or possibly related) serious adverse effect in each group, comparing these proportions using a Chi^2 test or Fisher's Exact test, as indicated. Specifically, we will report the number (%) of participants who: i) discontinued due to a related adverse event; ii) had 1 or more serious related adverse events (listing the type of serious related event and the number [%] for each type); and iii) had 1 or more non-serious related adverse events (listing the type of non-serious related events and the number [%] for each type).

Health economic analysis:

A health economic analysis will be undertaken. This analysis may be reported in a separate manuscript to the primary and secondary trial outcomes. A cost utility analysis will be undertaken from the perspective of the health system. The main outcome measure will be quality adjusted life years (QALYs) gained in the intervention group over the 52 week follow-up period as compared with QALYs gained in the control group. To calculate the QALYs gained, the responses to the EQ-5D-5L will be converted into a utility score anchored on a scale from 0 (indicating a health state equivalent to being dead) and 1 (indicating full health) using available preference weights from an Australian general population sample.[57] These utility scores will be converted into QALYs gained for each individual participant by combining utility scores with the information about the time the participant spend experiencing that utility using area under the curve methods. We will also utilise the primary clinical outcomes measures from the trial (Severity of overall knee symptoms of pain, stiffness and function [via WOMAC Total score]; and Physical activity levels [via average daily step count from wrist-worn accelerometry] as secondary outcome measures for assessing cost effectiveness. The incremental cost effectiveness ratios (ICER) comparing the costs for an improvement in the outcome (i.e., QALYs) in the intervention group as compared to the control group will be presented and their associated confidence intervals estimated. Cost effectiveness acceptability curves for varying threshold values of cost effectiveness will also be presented. An assessment of the sensitivity of the results to

variation in measured resource use effectiveness and/or unit costs will be undertaken using appropriate one-way and multi-way sensitivity analysis.[87,88]

Additional future analysis

A separate mediation analysis will be undertaken and reported in a separate manuscript to the primary and secondary trial outcomes. A full statistical analysis plan for the causal mediation analysis[89] will be developed separately to the current SAP and will be published online prior to analysis/unblinding.

Changes from the original published protocol

The following changes have been made in this SAP which were not present in the original published protocol:

- Physical activity analyses (sedentary, light, moderate/vigorous) have been changed from using Troiano et al.[34,35] cut-points to Hildebrand et al.[49] cut-points because the latter cut-points are specific for Actigraph accelerometry devices worn on the non-dominant wrist (i.e., consistent with trial data collection procedures).
- Consideration of physical activity intensity outcomes (minutes of sedentary, light, and moderate-vigorous activity) using Smuck et al cut-points,[36] and physical activity volume outcomes (intensity x time) using Smuck et al[36] and Hildebrand et al cut-points[49] as exploratory outcomes instead of secondary outcomes.
- All analyses using Smuck et al.[36] cut-points will be published as their own exploratory manuscript (separate from the trial manuscript). The Smuck et al.[36] cut-points were derived for hip-worn devices, and the conversion of cut-points between different wear locations is unknown (under investigation); this currently prevents analysis of our data attained from a wrist-worn wear location using these cut-points.
- Inclusion of 24-hour time use compositional data (expressed as isometric log ratios)[37,38] from wrist-worn accelerometry as an exploratory outcome. This will be published in an exploratory manuscript (separate from the trial manuscript).
- Inclusion of daily sleep time (minutes) identified via the Van Hees et al[39] sleep detection algorithm (guided by self-reported sleep logs) as an exploratory outcome.
- Inclusion of the proportion of people who meet or exceed the MCID for primary outcomes as a secondary outcome.
- Inclusion of pain self-efficacy as a potential moderator variable in subgroup analyses.
- Use of the FreKAQ as the measure of body perception (instead of tactile acuity or implicit motor imagery) as a potential moderator variable in subgroup analyses.
- Assessment of blinding status of the outcome assessors (collected the in-person lab measures at baseline and 12 weeks) as included in the original published protocol will now be included in the mediation analysis statistical analysis plan (i.e., lab measures are hypothesised mediators).
- Addition of the Bang Blinding Index to evaluate participant blinding.
- The proportion of therapists who thought they delivered the intervention of interest of the trial will be calculated (instead of successful blinding), given that the PSE-

informed care therapists were unavoidably aware that they were providing the intervention of interest. Thus, this will reflect a measure of therapist treatment expectancy rather than true blinding. Similar therapy treatment expectancy will be considered present if ≥80% of therapists in the control group were unsure or incorrectly thought they provided the intervention of interest of the trial.

- The original protocol identified that the health economic analysis will be published in a separate manuscript, with its own SAP developed. We have included the health economic analysis details here as this analysis may be published with that of the primary and secondary outcomes (dependent upon request of the journal).

Funding sources

This work was funded by the National Health & Medical Research Council (NHMRC) of Australia (ID1161634).

References:

1. March LM, Bagga H. Epidemiology of osteoarthritis in Australia. *Med J Aust.* 2004;180:S6-10.
2. National Institute for Health and Care Excellence. Osteoarthritis: care and management. Retrieved from: https://www.nice.org.uk/guidance/cg177/resources/osteoarthritis-care-and-management-pdf-35109757272517. Published 2020. Accessed.
3. Bennell KL, Hinman RS. A review of the clinical evidence for exercise in osteoarthritis of the hip and knee. *J Sci Med Sport.* 2011;14:4-9.
4. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews.* 2015(Issue 1. Art. No.: CD004576).
5. Wallis JA, Webster KE, Levinger P, Taylor NF. What proportion of people with hip and knee OA meet physical activity guidelines? A systematic review and meta-analysis. *Osteoarthr Cartilage.* 2013;21:1648-1659.
6. Williamson W, Kluzek S, Roberts N, et al. Behavioural physical activity interventions in participants with lower-limb OA: a systematic review and meta-analysis. *BMJ Open.* 2015;5:e007642.
7. Mazieres B, Thevenon A, Coudeyre E, Chevalier X, Revel M, Rannou F. Adherence to, and results of, physical therapy programs in patients with hip or knee OA. Development of French clinical practice guidelines. *Joint Bone Spine.* 2008;75.
8. Dobson F, Bennell KL, French SD, et al. Barriers and facilitators to exercise participation in people with hip and/or knee OA: Synthesis of the literature using behavior change theory. *Am J Phys Med Rehabil.* 2016;95:372-389.
9. Wilcox S, Der Ananian C, Abbott J, et al. Perceived exercise barriers, enablers and benefits among exercising and nonexercising adults with arthritis: results from a qualitative study. *Arthritis & Rheum.* 2006;55:616-627.
10. Bunzli S, O'Brien P, Ayton D, et al. Misconceptions and the Acceptance of Evidence-based Nonsurgical Interventions for Knee Osteoarthritis. A Qualitative Study. *Clin Orthop Relat Res.* 2019;477(9):1975-1983.
11. Pouli N, Nair RD, Lincoln NB, Walsh D. The experience of living with knee OA: exploring illness and treatment beliefs through thematic analysis. *Disabil Rehabil.* 2014;36:600-607.

12. Wallis JA, Taylor NF. Pre-operative interventions (non-surgical and non-pharmacological) for patients with hip or knee OA awaiting joint replacement surgery - a systematic review and meta-analysis. *Osteoarthr Cartilage.* 2011;19:1381-1395.
13. Larsson C, Ekvall Hansoon E, Sundquist K, Jakobsson U. Impact of pain characteristics and fear-avoidance beliefs on physical activity levels among older adults with chronic pain: a population-based, longitudinal study. *BMC Geriatrics.* 2016;16:50.
14. Quicke JG, Foster NE, Ogollah RO, Croft PR, Holden MA. The relationship between attitudes, beliefs and physical activity in older adults with knee pain: secondary analysis of a randomised controlled trial. *Arthritis Care Res.* 2017;69:1192-2000.
15. Moseley GL, Butler DS. Fifteen years of Explaining Pain: The past, present, and future. *J Pain.* 2015;16:807-813.
16. Smith BE, Moffatt F, Hendrick P, et al. Barriers and facilitators of loaded self-managed exercises and physical activity in people with patellofemoral pain: understanding the feasibility of delivering a multicentred randomised controlled trial, a UK qualitative study. *BMJ Open.* 2019;9(6):e023805.
17. Gay C, Chabaud A, Guilley E, Coudeyre E. Education patients about the benefits of physical activity and exercise for their hip and knee osteoarthritis. Systematic literature review. *Ann Phys Rehabil Med.* 2016;59:174-183.
18. Focht BC. Effectiveness of exercise interventions in reducing pain symptoms among older adults with knee osteoarthritis: A review. *J Aging Phys Act.* 2006;14:212-235.
19. Edwards RR, Dolman AJ, Martel MO, et al. Variability in conditioned pain modulation predicts response to NSAID treatment in patients with knee osteoarthritis. *BMC Musculoskelet Disord.* 2016;17:284-284.
20. Moseley GL. Evidence for a direct relationship between cognitive and physical change during an education intervention in people with chronic low back pain. *Eur J Pain.* 2004;8:-39-45.
21. Moseley GL, Nicholas MK, Hodges PW. A randomized controlled trial of intensive neurophysiology education in chronic low back pain. *Clin J Pain.* 2004;20:324-330.
22. Moseley GL. Combined physiotherapy and education is efficacious for chronic low back pain. *Physiother.* 2002;48:297-302.
23. Moseley GL. Joining forces - combining cognition-targeted motor control training with group or individual pain physiology education: A successful treatment for chronic low back pain. *J Man Manip Therapeutics.* 2003;11:88-94.
24. Ryan CG, Gray HG, Newton M, Granat MH. Pain biology education and exercise classes compared to pain biology education alone for individuals with chronic low back pain: A pilot randomised controlled trial. *Man Ther.* 2010;15:382-387.
25. Tegner H, Frederiksen P, Esbensen BA, Juhl C. Neurophysiological Pain Education for Patients With Chronic Low Back Pain: A Systematic Review and Meta-Analysis. *Clin J Pain.* 2018;34(8):778-786.
26. Wood L, Hendrick PA. A systematic review and meta-analysis of pain neuroscience education for chronic low back pain: Short-and long-term outcomes of pain and disability. *Eur J Pain.* 2019;23(2):234-249.
27. Liu Y, Ding W, Wang HL, et al. Gut microbiota and obesity-associated osteoarthritis. *Osteoarthritis Cartilage.* 2019;27(9):1257-1265.
28. Morris JL, Letson HL, Gillman R, et al. The CNS theory of osteoarthritis: Opportunities beyond the joint. *Semin Arthritis Rheum.* 2019;49(3):331-336.
29. Helmark IC, Mikkelsen UR, Børglum J, et al. Exercise increases interleukin-10 levels both intraarticularly and peri-synovially in patients with knee osteoarthritis: a randomized controlled trial. *Arthritis Res Ther.* 2010;12(4):R126-R126.

30. Roos EM, Dahlberg L. Positive effects of moderate exercise on glycosaminoglycan content in knee cartilage: a four-month, randomized, controlled trial in patients at risk of osteoarthritis. *Arthritis Rheum.* 2005;52(11):3507-3514.
31. Runhaar J, Beavers DP, Miller GD, et al. Inflammatory cytokines mediate the effects of diet and exercise on pain and function in knee osteoarthritis independent of BMI. *Osteoarthritis Cartilage.* 2019;27(8):1118-1123.
32. Butler DS, Moseley GL, Stanton TR. *The Explain Pain Handbook: Knee osteoarthritis.* Adelaide, South Australia: NOIgroup publications; NOIgroup Australasia Pty Ltd.; 2023.
33. Stanton TR, Karran EL, Butler DS, et al. A pain science education and walking program to increase physical activity in people with symptomatic knee osteoarthritis: a feasibility study. *Pain Rep.* 2020;5(5):e830.
34. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181-188.
35. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine.* 2014;48(13):1019-1023.
36. Smuck M, Tomkins-Lane C, Ith MA, Jarosz R, Kao MJ. Physical performance analysis: A new approach to assessing free-living physical activity in musculoskeletal pain and mobility-limited populations. *PLoS One.* 2017;12(2):e0172804.
37. Dumuid D, Pedišić Ž, Stanford TE, et al. The compositional isotemporal substitution model: A method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. *Stat Methods Med Res.* 2019;28(3):846-857.
38. Dumuid D, Stanford TE, Martin-Fernández JA, et al. Compositional data analysis for physical activity, sedentary time and sleep research. *Stat Methods Med Res.* 2018;27(12):3726-3738.
39. van Hees VT, Sabia S, Anderson KN, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. *PLOS ONE.* 2015;10(11):e0142533.
40. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology.* 2002;2(1):8.
41. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ.* 1998;316(7139):1236-1238.
42. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "p < 0.05". *The American Statistician.* 2019;73:1-19.
43. Greenland S. Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. *The American Statistician.* 2019;73(sup1):106-114.
44. Yelland LN, Sullivan TR, Voysey M, Lee KJ, Cook JA, Forbes AB. Applying the intention-to-treat principle in practice: Guidance on handling randomisation errors. *Clin Trials.* 2015;12(4):418-423.
45. 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.
46. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. *J Clin Epidemiol.* 2005;58(6):595-602.
47. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient

relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15(12):1833-1840.
48. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain.* 2001;94(2):149-158.
49. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. *Scandinavian Journal of Medicine & Science in Sports.* 2017;27(12):1814-1823.
50. Cella D, Choi SW, Condon DM, et al. PROMIS(®) Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value Health.* 2019;22(5):537-544.
51. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385-396.
52. Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). *Journal of Health Psychology.* 2013;18(12):1617-1628.
53. Shelby RA, Somers TJ, Keefe FJ, et al. Brief fear of movement scale for OA. *Arthritis Care Res.* 2012;64:862-871.
54. McWilliams LA, Kowal J, Wilson KG. Development and evaluation of short forms of the Pain Catastrophizing Scale and the Pain Self-efficacy Questionnaire. *European Journal of Pain.* 2015;19(9):1342-1349.
55. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. *Eur J Pain.* 2007;11(2):153-163.
56. Pulling BW, Braithwaite FA, Butler DS, et al. Item development and pre-testing of an Osteoarthritis Conceptualisation Questionnaire to assess knowledge and beliefs in people with knee pain. *PLOS ONE.* 2023;18(9):e0286114.
57. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research.* 2011;20(10):1727-1736.
58. Nishigami T, Mibu A, Tanaka K, et al. Development and psychometric properties of knee-specific body-perception questionnaire in people with knee osteoarthritis: The Fremantle Knee Awareness Questionnaire. *PLoS One.* 2017;12(6):e0179225.
59. ten Klooster PM, Drossaers-Bakker KW, Taal E, van de Laar MA. Patient-perceived satisfactory improvement (PPSI): interpreting meaningful change in pain from the patient's perspective. *Pain.* 2006;121(1-2):151-157.
60. Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. *J Orthop Sports Phys Ther.* 2012;42(8):716-723.
61. White DK, Tudor-Locke C, Zhang Y, et al. Daily walking and the risk of incident functional limitation in knee osteoarthritis: an observational study. *Arthritis Care Res (Hoboken).* 2014;66(9):1328-1336.
62. Arel-Bundock V, Greifer N, Heiss A. How to interpret statistical models using marginaleffects in R and Python. *Journal of Statistical Software.* 2024.
63. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology.* 2017;17(1):162.
64. Jolani S. Hierarchical imputation of systematically and sporadically missing data: An approximate Bayesian approach using chained equations. *Biometrical Journal.* 2018;60(2):333-351.
65. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software.* 2011;45(3):1 - 67.

66. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology.* 2019;110:63-73.
67. Allaire J, Teague C, Scheidegger C, Xie Y, Dervieux C. Quarto (Version 1.4). In: Computer Software; 2024.
68. Knuth DE. Literate Programming. *The Computer Journal.* 1984;27(2):97-111.
69. Angrist JD, Imbens GW. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association.* 1995;90(430):431-442.
70. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health.* 2000;21:121-145.
71. Stuart EA, Perry DF, Le HN, Ialongo NS. Estimating intervention effects of prevention programs: accounting for noncompliance. *Prev Sci.* 2008;9(4):288-298.
72. Connell AM. Employing Complier Average Causal Effect Analytic Methods to Examine Effects of Randomized Encouragement Trials. *The American Journal of Drug and Alcohol Abuse.* 2009;35(4):253-259.
73. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med.* 2017;377(14):1391-1398.
74. Murray EJ, Swanson SA, Hernan MA. Guidlines for Estimating Causal Effects in Pragmatic Randomised Tirals (Draft for Public Comment, v2). *arXiv.* 2019;35.
75. Tardif H, Allingham SF, Rahman M, Daly A. Patterns of patient outcomes following specialist pain management in Australasia: a latent class analysis using the electronic Persistent Pain Outcomes Collaboration database. *Pain.* 2023;164(5):967-976.
76. Tanaka S, Nishigami T, Wand BM, et al. Identifying participants with knee osteoarthritis likely to benefit from physical therapy education and exercise: A hypothesis-generating study. *Eur J Pain.* 2021;25(2):485-496.
77. Pintrich PR, Smith DAF, Duncan T, et a. *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, Michigan: National Center for Research to Improve Postsecondary Teaching and Leanring.; 1991.
78. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry.* 2000;31(2):73-86.
79. Rhodes RE, Hunt Matheson D, Mark R. Evaluation of Social Cognitive Scaling Response Options in the Physical Activity Domain. *Measurement in Physical Education and Exercise Science.* 2010;14(3):137-150.
80. Bang H, Flaherty SP, Kolahi J, Park J. Blinding assessment in clinical trials: A review of statistical methods and a proposal of blinding assessment protocol. *Clinical Research and Regulatory Affairs.* 2010;27(2):42-51.
81. Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Control Clin Trials.* 2004;25(2):143-156.
82. Moroz A, Freed B, Tiedemann L, Bang H, Howell M, Park JJ. Blinding measured: a systematic review of randomized controlled trials of acupuncture. *Evid Based Complement Alternat Med.* 2013;2013:708251.
83. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-399.
84. Coertjens L, Donche V, De Maeyer S, Vanthournout G, Van Petegem P. To what degree does the missing-data technique influence the estimated growth in learning strategies over time? A tutorial example of sensitivity analysis for longitudinal data. *PLoS One.* 2017;12(9):e0182615.

85. 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.
86. Ashbeck EL, Bell ML. Single time point comparisons in longitudinal randomized controlled trials: power and bias in the presence of missing data. *BMC Medical Research Methodology.* 2016;16(1):43.
87. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation.* Oxford University Press; 2016.
88. Drummond M, Sculpher M, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* 4th ed. Oxford: Oxford University Press; 2015.
89. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods.* 2010;15:309-334.

Linear mixed effects model formulation: primary analysis

EPIPHAKNEE | subSAP_PA_v10 | 2024-Jul-05

Table of contents

1 Document revision history

Purpose	An auxiliary technical document to the EPIPHAKNEE statistical analysis plan (SAP) to provide further mathematical detail and principles of the intended primary analysis linear mixed effects modelling
Identifier	subSAP_PA_v10
Version	v1.0
Parent document	EPIPHAKNEE SAP [v1.0]
Prepared by	Ty Stanford
Signature	
Date	2024-Jul-05

2 Overview of maximally complex LMM Model under consideration

As outlined in the parent document (SAP [v1.0]) the primary analyses will report "marginal means" and confidence intervals based on the linear mixed effect modelling. This document outlines further mathematical detail and principles of the intended primary analysis linear mixed modelling.

2.1 Primary analysis

The random variable Y_{kij} denotes the unrealised outcome for participant $i = 1, 2, \ldots, n$ under the care of physiotherapist $k = 1, 2, \ldots, K$ at follow-up time $j = 0, 1, 2, 3$; corresponding to trial follow-up times: baseline, 12 weeks, 26 weeks and 52 weeks, respectively. We propose the high-level regression model structure as follows,

$$Y_{kij} = \eta_i + \zeta_i^{(j)} + \psi_{ij} + V_{kij}, \tag{1}$$

where Greek characters (η, ζ, ψ) represent fixed effects and Latin symbols (V) represent random effects; and

- η_i contains the (fixed) effects and non-time-varying predictor variables, including the fixed intercept, where participant i's characteristics are constant over follow-ups $j = 0, 1, 2, 3$;
- $\zeta_i^{(j)}$ contains the (fixed) effects and potentially time-varying predictor variables where participant i's characteristics potentially vary over follow-ups $j = 0, 1, 2, 3$;
- ψ_{ij} contains the intervention arm by time interaction and main (fixed) effects for participant i at follow-up time j (the main predictors of interest for the trial);
- V_{kij} are the multilevel (random) effects, which includes a random participant i intercept (and possibly correlated random slope or time-specific random intercept depending on whether time-point is treated as a continuous or categorical variable, respectively), random intercept for physiotherapist k, and the residual error for participant i under the care of physiotherapist k at follow-up time-point j.

The treatment of the follow-up time variable will be a data and model diagnostic informed choice (continuous with or without transformation or categorical variable).

However, the terms η_i (non-time-varying fixed effects) and $\zeta_i^{(j)}$ (fixed effects and potentially time-varying predictor variables) are to be fit irrespective of the follow-up time-point operationalisation (variable type selection). These sets of terms are outlined in Section 3, prior to outlining the terms which are dependent on the follow-up time-point operationalisation (Section 4.1 and Section 4.2).

2.2 Sensitivity analysis: potential COVID-19 restriction effects

The sensitivity analysis to estimate COVID-19 restriction effects, takes the same form as Equation 1 however and additional set of terms (ξ_{ij}) are added:

$$Y_{kij} = \eta_i + \zeta_i^{(j)} + \psi_{ij} + \xi_{ij} + V_{kij}. \tag{2}$$

ξ_{ij} contains the potential COVID-19 restriction (fixed) effects on the group by time interaction for participant i at follow-up time-point j.

The terms ψ_{ij}, ξ_{ij} and V_{kij} will require different formulations depending on the follow-up time-point variable type choice. Section 4.1 and Section 4.2 outline these scenarios and the resultant forms of ψ_{ij}, ξ_{ij} and V_{kij}.

3 Model terms independent of follow-up time operationalisation

3.1 Fixed effects with constant patient characteristics (η_i)

The η_i set of terms can be more fully expressed as

$$\eta_i = \alpha_0 + \sum_{\ell=1}^{p} \alpha_\ell x_{\ell i}$$

where

- $x_{\ell i}$ are the p ($\ell = 1, 2, \ldots, p$) non-time varying variables/characteristics (either continuous or indicator variables representing non-reference categorical levels) of patient i outlined in §7 (*Analysis*) of the SAP [v1.0]; and
- $\alpha_0, \alpha_1, \ldots, \alpha_p$ are the p fixed coefficients to be estimated by the model.

3.2 Fixed effects with potentially varying patient characteristics ($\zeta_i^{(j)}$)

The ζ_{ij} set of terms can be more fully expressed as

$$\zeta_i^{(j)} = \sum_{r=1}^{q} \delta_r x_{ri}^{(j)}$$

where

- $x_{ri}^{(j)}$ are the q ($r = 1, 2, \ldots, q$) potentially time-varying variables/characteristics (either continuous or indicator variables representing non-reference categorical levels) of patient i at each follow-up time-point $j(= 0, 1, 2, 3)$, outlined in §7 (*Analysis*) of the SAP [v1.0]; and
- $\delta_1, \ldots, \delta_q$ are the q fixed coefficients to be estimated by the model.

4 Operationalisation of the follow-up time variable

The time variable will initially be attempted to be operationalised as a continuous variable so that a random slopes for patients (and correlated with the corresponding random intercepts) can be fit and tested for significance via log-likelihood ratio tests (LRTs).

Model diagnostics (importantly, linearity between conditional residuals vs predicted values – by time-point and group) will inform the appropriate operationalisation of the time variable.

The order of preferred treatment of the follow-up time variable is as follows:

1. continuous without transformation,
2. continuous with transformation (please see Figure 1 for transformations considered, in order), and finally
3. categorical.

Figure 1 outlines the simple transformations of (continuous) time that will be considered, in order left to right, with the final time variable considered reverting to a flexible 4-level categorical variable. If residual linearity is achieved using an untransformed or (non-quadratic) transformed continuous time variable but there exists apparent residual heteroscedasticity over time, two simple transformations of the strictly positive outcome variable(s) will be considered to create variance stabilisation. In the case variance stabilisation cannot be achieved using a (un)transformed continuous time variable, or residual linearity cannot achieved using a (un)transformed continuous time variable; a flexible categorical follow-up time variable will be used.

The previously mentioned random slopes will not be able to be if the time variable is categorical, however residual-side heterogeneous and unstructured correlated random errors can be fit to account for residual heteroscedasticity (and correlation) across time.
Random effect parsimony will be tested via a LRT of competing models.

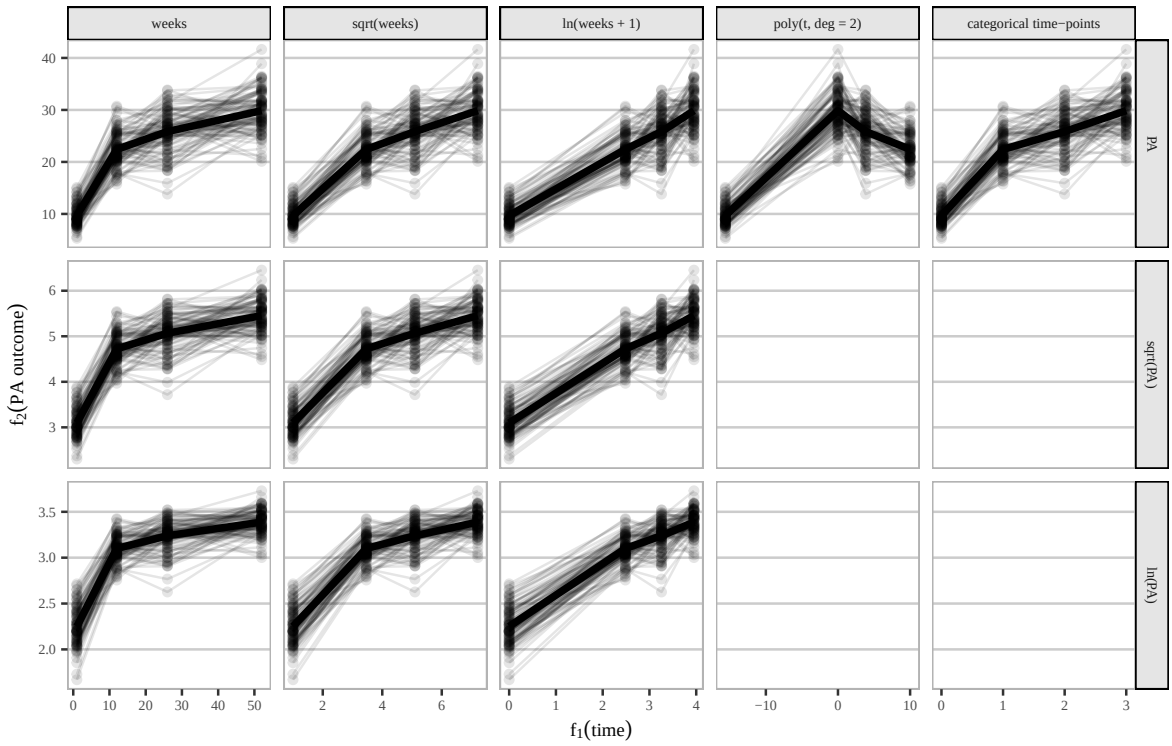

Figure 1: Conceptualisation of possible follow-up time transformations to linearise its relationship with the outcomes (if required) in addition to possible outcome transformations to create variance stabilisation in the case of residual variance being a function of time – made more likely as the outcome variables can only take non-negative values. Transformations are limited to simple monotonic functions for understandability and ease of reporting.

The entire model fitting, selection and diagnostic processes will be made available in supplementary files using literate programming principles for reproducibility and transparency.

4.1 Case A: Continuous follow-up time variable

In the case a transformation of the follow-up time variable is satisfactory viable (pre-specified in Figure 1; satisfactory based on model fit diagnostics, Section 4), we will define the time-point variables for participant i as

$$t_{ij} = f(w_{ij})$$

where $j = 0, 1, 2, 3$ as before with corresponding follow-up times as weeks where $\{w_{i0} = 0, w_{i1} = 12, w_{i2} = 26, w_{i3} = 52\}$, and $f(.)$ is the transformation function (including $f(a) = a$ and the "untransformed" w_{ij} value).

4.1.1 Fixed effects of interest for the trial: group $\times$ time (ψ_{ij})

In regards to the intervention arm/group, define the variable g_i that takes the value 0 if participant i is in the control group and 1 if participant i is in the PSE-informed care intervention.

Subsequently, ψ_{ij} from Equation 1 will contain both main effect contrasts between intervention groups at each follow-up time-point *and* additional interaction effects of PSE-informed care intervention with follow-up time-point:

$$\beta_1\ g_i + \beta_2\ t_{ij} + \beta_3\ (g_i \times t_{ij})$$

where

- β_1, β_2 and β_3 are fixed coefficients to be estimated by the model.

Note that $\hat{\beta}_3$ is the estimated additional over-time change (slope) attributed to the PSE-informed care intervention, when compared to the common estimated slope $\hat{\beta}_2$ common to both intervention arms.

In the case of a degree=2 polynomial being chosen as the time-point transformation, ψ_{ij} from Equation 1 will take the form

$$\beta_1\ g_i + \beta_2\ t_{ij} + \beta_3\ t_{ij}^2 + \beta_4\ (g_i \times t_{ij}) + \beta_5\ (g_i \times t_{ij}^2)$$

where $t_{ij} = j - 1$ for ease of interpretation of the linear and quadratic time coefficients.

Note, in the specific case of polynomial time predictors, model reduction will be considered in the following order:

- the removal of intervention arm/group specific quadratic time effect ($\beta_5 = 0$).

4.1.2 Random effects (V_{kij})

The random elements of Equation 1 (and Equation 2), namely V_{kij}, will initially be fit with the following random effects

$$u_k + b_{0i} + b_{1i} \times t_{ij} + e_{kij}$$

where

- u_k is a random intercept associated with physiotherapist k, with $u_k \sim N(0, \sigma_u^2)$;
- b_{0i} is a random intercept for participant i, and b_{1i} is a random slope for participant i with

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N(\mathbf{0}, \Sigma_b), \text{ and}$$

$$\Sigma_b = \begin{bmatrix} \sigma_0^2 & \rho_{01}\sigma_0\sigma_1 \\ \rho_{01}\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix}; \text{ and}$$

- e_{kij} is the random residual error for participant i under the care of physiotherapist k at follow-up time-point j, with $e_{kij} \sim N(0, \sigma_e^2)$; and
- u_k, $\boldsymbol{b}_i$ and e_{kij} assumed to be independently distributed, i.e.,

$$Var\left(\begin{bmatrix} u_k \\ \boldsymbol{b}_i \\ e_{kij} \end{bmatrix}\right) = \begin{bmatrix} \sigma_u^2 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{2\times 1} & \Sigma_b & \mathbf{0}_{2\times 1} \\ 0 & \mathbf{0}_{1\times 2} & \sigma_e^2 \end{bmatrix}.$$

More parsimonious random effect structures will be investigated where:

- $\rho_{01} = 0$ (independent participant random intercepts and slopes), and
- σ_1 (no random slope variability - only participant random intercepts required).

4.1.3 COVID-19 restriction moderation on the group $\times$ time interaction (ξ_{ij})

The terms in ξ_{ij} from Equation 2 contain the same predictors as in ψ_{ij}, however they are crossed with the time specific ($j = 0, 1, 2, 3$) COVID-19 restriction variable for participant i that we name c_{ij}. We assume $c_{ij} = 0$ means no COVID-19 restrictions and takes the value 1 for any COVID-19 restrictions. Therefore, ξ_{ij} can be written

$$c_{ij} \times \{\beta_4\, g_i + \beta_5\, t_{ij} + \beta_6\, (g_i \times t_{ij})\}$$

where

- $\beta_4, \beta_5, \beta_6$ are fixed coefficients to be estimated by the model.

Model reduction will consider in the following order::

- the removal of intervention arm/group specific COVID-19 restriction effects ($\beta_6 = 0$),
- the removal of time specific COVID-19 restriction effects ($\beta_5 = 0$), and finally
- any constant COVID-19 restriction effects ($\beta_4 = 0$).

Note that in the case of a degree=2 polynomial being chosen as the time-point transformation, the above still will be used for the terms in ξ_{ij} (NOT considering quadratic time COVID-19 interactions), but the indexes of the betas will be different. Model reduction will prioritise removal of the quadratic terms if possible.

4.2 Case B: Categorical follow-up time variable

In the case the follow-up time variable is categorical, we will define the indicator time-point variables for participant i as

$$t_{ij'} = \mathrm{I}\left(j' = j\right)$$

where $j' = 1, 2, 3$ (as they will be associated with contrasts to baseline) and $j = 0, 1, 2, 3$ as before, and I(.) is the indicator function that only takes values 0 (when the argument is false) or 1 (when the argument is true).

For example, for participant $i = 5$ at follow-up $j = 2$, the corresponding indicator variables are

- $t_{51} = 0$,
- $t_{52} = 1$, and
- $t_{53} = 0$.

4.2.1 Fixed effects of interest for the trial: group $\times$ time (ψ_{ij})

In regards to the intervention arm/group, define the variable g_i that takes the value 0 if participant i is in the control group and 1 if participant i is in the PSE-informed care intervention.

Subsequently, ψ_{ij} from Equation 1 will contain both main effect contrasts between intervention groups at each follow-up time-point *and* additional interaction effects of PSE-informed care intervention with follow-up time-point:

$$\left(\beta_1\, t_{i1} + \beta_2\, t_{i2} + \beta_3\, t_{i3}\right) + g_i \times \left(\beta_4\, t_{i1} + \beta_5\, t_{i2} + \beta_6\, t_{i3}\right)$$

where

- $\beta_1, \beta_2, \ldots, \beta_6$ are fixed coefficients to be estimated by the model.

Note that $\hat{\beta}_6$ is the specific estimated contrast point estimat between the mean adjusted change from baseline to the 52 week follow-up between the PSE-informed care intervention and the control.

4.2.2 Random effects (V_{kij})

The random elements of Equation 1 (and Equation 2), namely V_{kij}, will initially be fit with the following random effects

$$u_k + b_i + e_{kij}$$

where

- u_k is a random intercept associated with physiotherapist k, with $u_k \sim N(0, \sigma_u^2)$;
- b_i is a random intercept for participant i, with $b_i \sim N(0, \sigma_b^2)$;
- e_{kij} is the random residual error for participant i under the care of physiotherapist k at follow-up time-point j, with

$$\boldsymbol{e}_{ki} = \begin{bmatrix} e_{ki0} \\ e_{ki1} \\ e_{ki2} \\ e_{ki3} \end{bmatrix} \sim N(\mathbf{0}, \Sigma), \text{ and}$$

$$\Sigma = \begin{bmatrix} \sigma_0^2 & \rho_{01}\sigma_0\sigma_1 & \rho_{02}\sigma_0\sigma_2 & \rho_{03}\sigma_0\sigma_3 \\ \rho_{01}\sigma_0\sigma_1 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{02}\sigma_0\sigma_2 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{03}\sigma_0\sigma_3 & \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}; \text{ and}$$

- u_k, b_i and $\boldsymbol{e}_{ki}$ assumed to be independently distributed, i.e.,

$$Var\left(\begin{bmatrix} u_k \\ b_i \\ \boldsymbol{e}_{ki} \end{bmatrix}\right) = \begin{bmatrix} \sigma_u^2 & 0 & \mathbf{0}_{1\times 4} \\ 0 & \sigma_b^2 & \mathbf{0}_{1\times 4} \\ \mathbf{0}_{4\times 1} & \mathbf{0}_{4\times 1} & \Sigma \end{bmatrix}.$$

More parsimonious random effect structures will be investigated where:

- $\rho_{01} = \rho_{02} = \rho_{03} = \rho_{12} = \rho_{13} = \rho_{23} = 0$, and
- $\sigma_0 = \sigma_1 = \sigma_2 = \sigma_3$.

4.2.3 COVID-19 restriction moderation on the group $\times$ time interaction (ξ_{ij})

The terms in ξ_{ij} from Equation 2 contain the same predictors as in ψ_{ij}, however they are crossed with the time specific ($j = 0, 1, 2, 3$) COVID-19 restriction variable for participant i that we name c_{ij}. We assume $c_{ij} = 0$ means no COVID-19 restrictions and takes the value 1 for any COVID-19 restrictions. Therefore, ξ_{ij} can be written

$$c_{ij} \times \{(\beta_7 + \beta_8\, t_{i1} + \beta_9\, t_{i2} + \beta_{10}\, t_{i3}) + g_i \times (\beta_{11} + \beta_{12}\, t_{i1} + \beta_{13}\, t_{i2} + \beta_{14}\, t_{i3})\}$$

where

- $\beta_7, \beta_8, \ldots, \beta_{14}$ are fixed coefficients to be estimated by the model.

Model reduction will consider:

- the removal of intervention arm/group specific COVID-19 restriction effects ($\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$),
- the removal of time specific COVID-19 restriction effects ($\beta_8 = \beta_9 = \beta_{10} = 0$), and finally
- any constant COVID-19 restriction effects ($\beta_7 = 0$).