Statistical Analysis Plan (SAP)

Trial: Footwear for self-managing knee osteoarthritis symptoms: the Footstep Trial

Contents

Section 1. Administrative Information

1. Title

Footwear for self-managing knee osteoarthritis symptoms: The Footstep Trial

2. Trial registration

Prospectively registered (Australian New Zealand Clinical Trials Registry Trial Id: ACTRN12617001098325, 28/07/2017)

3. SAP version

Version: 1.0 Date: 10/06/2020

4. Protocol Version

This document has been written based on information contained in the Footstep study protocol version 1.3 dated 12/06/19. The protocol was published as follows:

Paterson K.L., et al., Footwear for self-managing knee osteoarthritis symptoms: protocol for the Footstep randomised controlled trial. *BMC Musculoskelet Disord.* 2018;19(1):219. Published 2018 Jul 18. doi:10.1186/s12891-018-2144-1

5. SAP Revisions

Not applicable

6. Names and affiliations

Document prepared by Professor Rana Hinman, Mr Tim Wrigley, Dr Kade Paterson, Mr Ben Metcalf, Mrs Penny Campbell and Professor Kim Bennell, University of Melbourne and Dr Jessica Kasza, Monash University.

Emails: RS Hinman ranash@unimelb.edu.au; TV Wrigley timw@unimelb.edu.au; KL Paterson kade.paterson@unimelb.edu.au; BR Metcalf b.metcalf@unimelb.edu.au; PK Campbell penelope.campbell@unimelb.edu.au; KL Bennell k.bennell@unimelb.edu.au; J Kasza jessica.kasza@monash.edu;

Signatures:

Signature of senior statistician responsible (Dr Jessica Kasza): *Date: 10/6/2020*

Signature of chief investigator (Prof Rana Hinman): *Date: 10/6/2020*

Section 2: Introduction

7. Background and rationale

Osteoarthritis (OA) is the leading cause of musculoskeletal pain and disability in Australia and the medial knee compartment is most often affected. In 2012, 1.9 million Australians had OA,[1] and modelling data estimate a 58% increase in OA by 2032.[1] Knee OA is extremely debilitating. Pain is dominant, becoming persistent and more limiting as OA progresses. OA is the 11th highest contributor to global disability,[2] and arthritis is the 6th most-managed problem by Australian general practitioners.[3] Health expenditure on OA in Australia in 2012 was $3.75 billion,[1] with most costs related to conservative and surgical treatments, lost productivity and substantial loss of quality of life. Accordingly, the Australian Government has designated arthritis a National Health Priority and developed a National Action Plan to reduce disease burden.

As OA is a chronic disease with no cure, people with OA have little choice but to self-manage their condition. Accordingly, advice about self-management is the cornerstone of conservative treatment, along with exercise and weight control.[4, 5] As abnormal biomechanics are central to disease pathogenesis,[6, 7] clinical guidelines advocate clinicians provide advice on "appropriate" footwear as part of core treatment for knee OA.[4, 8] However, there is scant evidence from clinical trials to guide footwear choice for this patient group, or indeed, any population with chronic musculoskeletal pain. Guidelines suggest that shoes with thick shock-absorbing soles and arch supports are best for people with OA),[8] based largely on expert opinion. Due to the lack of robust clinical trials in this area, footwear trials have been identified as an OA research priority by both the European League Against Rheumatism [8] and the National Institute for Health and Care Excellence (UK).[4]

8. Objectives

Research hypothesis:

Primary alternative hypothesis: That flat flexible shoes will lead to significantly greater reductions in knee pain with walking and improvements in physical function at 6 months, compared to stable supportive shoes.

Secondary alternative hypothesis: That flat flexible shoes will have significantly greater benefits on other clinical outcomes (overall knee pain, global ratings of change, pain at other sites, health-related quality of life, physical activity levels) compared to stable supportive shoes at 6 months.

Study objective:

Primary objective: To determine if flat flexible shoes lead to significantly greater reductions in knee pain with walking, and improvements in physical function, compared to stable supportive shoes at six months.

Secondary objective: To determine if flat flexible shoes will have significantly greater benefits on other clinical outcomes (other parameters of knee pain, sport and recreational function, knee-related quality of life, global ratings of change, pain at other sites, health-related quality of life, physical activity levels) compared to stable supportive shoes at 6 months.

Section 3: Trial Methods

9. Trial design

A two-arm pragmatic, comparative effectiveness, parallel-group randomised controlled trial. Treatment allocation is a 1:1 ratio. Participants are randomized to either flat flexible shoes or stable supportive shoes.

10. Randomisation

Eligible participants will be randomised to receive either flat flexible or stable supportive shoes. The randomisation schedule will be prepared by the biostatistician (permuted block sizes 6 to 12) stratified by radiographic disease severity (Kellgren & Lawrence grade 3 or 4). The schedule will be stored on a password-protected website (REDCap) maintained by a researcher not involved in either participant recruitment or administration of primary/secondary outcome measures. Group allocation will be revealed by this same researcher after baseline primary/secondary outcomes have been completed.

11. Sample size

We aim to detect the minimal clinically important difference (MCID) on primary outcomes between groups (1.8 (out of 10) [9] for Numerical Rating Scale pain and 6 (out of 68) for Western Ontario & McMaster Universities OA Index function).[10] We assume between-subject standard deviations of 2.7 and 11.4, and baseline to 6-month correlations of 0.21 and 0.39 for pain and function respectively (data from our footwear trial in a similar sample).[11] Using analysis of covariance adjusted for baseline score, we need 46 per arm to achieve 90% power to detect MCID in pain and 65 per arm for function. Allowing for 20% attrition, we will recruit 82 people per arm (n=164).

12. Framework

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

13. Statistical Interim analyses and stopping guidance

Nil

14. Timing of final analysis

Final analysis will be performed after all (n=164) participants have reached the 6-month timepoint and completed assessments.

15. Timing of outcome assessments

Table 4.6 in the study protocol details the timing of outcome assessments, the majority of which occur at baseline and at 6-months.

Section 4: Statistical Principles

16. Level of statistical significance

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

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

Due to the expected correlation between the pain and function primary outcomes, we will not adjust for multiplicity.[12]

18. Confidence intervals to be reported

All confidence intervals will be 95% confidence intervals.

19. Adherence and Protocol Deviations

The primary analysis will be based on the principle of intention-to-treat, whereby participants are included in the groups to which they were originally assigned, regardless of their adherence to their assigned treatments. Any protocol deviations (if they occur), including errors applying inclusion/exclusion criteria and/or administration of the wrong intervention will be summarised in trial results (patient flow diagram/text) by treatment group. Randomisation errors resulting from these errors will be handled according to recommendations. [13]

Multiple measures of adherence are used in this trial (described in Table 4.6 of the protocol) and data from all measures will be reported using means, standard deviations and proportions (number and percentage) as appropriate for each treatment group.

In this study, participants are advised to wear their allocated shoes for at least 6 hours per day. Participants will be classified as adherent to the intervention (or not) based on the log-book data recording daily allocated shoe wear (hours per day). Participants complete one log-book per month for a 7 day period (snapshot), reporting the number of hours each day that they wore each of their two pairs of allocated shoes. The hours reported for each pair of shoes will be summed for each reported day to give a daily total. The daily total will be averaged over the 7 days for each of the six reporting periods (months). The mean (SD) hrs/day of wearing allocated shoes will be reported for each month of the six months across treatment groups. The number (%) of adherent (average daily shoe wear $\geq$ 6 hrs) and non-adherent (average daily shoe wear $<$ 6 hrs) participants will be reported for each treatment group at each month over the six months.

For each participant, average daily allocated shoe wear (hrs/day) over the total 6-month intervention period will be calculated as the average of the six reporting period averages. Participants who average at least 6 hrs/day of wearing their allocated trial shoes over the 6-month intervention period will be classified as adherent, and those who don't will be classified as non-adherent. The number (%) of adherent (average daily shoe wear $\geq$ 6 hrs) and non-adherent (average daily shoe wear $<$ 6 hrs) participants will be reported for each treatment group for the entire 6-month intervention.

If a participant does not provide all log books, average daily allocated shoe wear will be calculated using the available log books. If a participant does not provide any log books, single mean imputation by treatment group will be applied.

20. Analysis Populations

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 randomised, regardless of their adherence to their assigned treatments.

Section 5: Trial Population

21. Screening Data

Screening data will be collected and summarized. A CONSORT flow diagram will be used.[14] The following summaries will be presented in text and/or flow diagram: time frame for recruitment, the number of patients screened, the number of patients recruited, the number of screened patients not recruited, and the reasons for non-recruitment.

22. Eligibility

Trial inclusion and exclusion criteria are described in section 5.2 of the trial protocol. Reasons for exclusion will be summarized in the CONSORT [14] flow diagram.

23. Recruitment

A CONSORT flow diagram [14] will be used to describe the number of people enrolled, randomized, allocated to each treatment group, lost to follow up (including reasons) and analysed.

24. Withdrawal/follow-up

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

25. Baseline characteristics

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

- Age
- Gender
- Height, body mass, body mass index
- Radiographic disease severity using the Kellgren & Lawrence scale
- Duration of knee OA symptoms
- Current employment status
- Expectation of treatment outcome (before and after shoe allocation)
- Arthritis Self Efficacy Scale
- Co-intervention use
- Objective measures of foot posture including:
    - Foot posture index
    - Foot mobility magnitude
    - Navicular drop
    - Footwear characteristics of the most commonly worn pair of own shoes, reported as the number and proportion of participants who currently use flat flexible shoes, stable supportive shoes and/or shoes of mixed features.

Although the protocol describes baseline measures of in-shoe plantar pressures while walking, this data will not be described in the trial report as it is complex biomechanical data that is not measured in typical clinical practice (thus does not influence interpretation of trial findings) and is not going to be used in the planned moderation analyses. Instead, this data will be reported in a separate cross-sectional paper.

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 and compare these characteristics between two groups: those participants who provide both primary outcomes at 6 months, and those participants who are missing one or both primary outcomes. 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

Co-primary outcomes:

- Change in severity of knee pain during walking: Average overall knee pain on walking in the past week is self-assessed using a 11-point numeric rating scale (NRS) with terminal descriptors of 'no pain' (score 0) and 'worst pain possible' (score 10). Change score at 6 months will be calculated as baseline minus follow-up.

- Change in physical function subscale of the WOMAC: Limitations with physical functioning will be measured by the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index (Likert version 3.1). The WOMAC is a self-report, disease-specific instrument which has established validity, reliability and responsiveness in an extensive range of OA studies. The WOMAC physical function subscale contains 17 questions regarding knee function, with Likert response options ranging from 0 (no dysfunction) to 4 (extreme dysfunction). Total score ranges from 0 to 68, with higher scores indicating worse function. WOMAC scores will be extracted from the Knee Injury and Osteoarthritis Outcome Score (KOOS) questionnaire, which contains the WOMAC questions. Change score at 6 months will be calculated as baseline minus follow-up.

Secondary outcome:

- Change in each of pain, function in sport and recreation, knee-related quality of life and patellofemoral pain and OA in the last week: The additional four subscales of the Knee Injury and Osteoarthritis Outcome Score (KOOS) are being measured including i) pain (9-items), ii) function in sport and recreation (5-items), iii) knee-related quality of life quality of life (4-items) and iv) patellofemoral pain and OA (11-items). Responses are provided on a 5-point scale. Scores will be calculated for each subscale and range from 0 to 100 with 0 indicating worst possible symptoms. Change scores at 6 months will be calculated as baseline minus follow-up.

- Change in average overall pain severity over past week at other sites: i) study knee, ii) contralateral knee, iii) ipsilateral hip, iv) contralateral hip, v) ipsilateral foot/ankle, vi) contralateral foot/ankle, vii) back will each be scored on an 11-point NRS for average overall pain in the last week. Scores range from 0 to 10; where 0=no pain and 10=worst pain possible. Change scores at 6 months will be calculated as baseline minus follow-up.

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

- Change in physical activity: The Physical Activity Scale for the Elderly is being used to assess physical activity over the previous week. This is a 10-item questionnaire which collects responses for the frequency, duration and intensity level of a range of activities typically chosen by older adults. Scores range 0 to >400 with higher scores indicating greater levels of physical activity. Change scores at 6 months will be calculated as baseline minus follow-up.

- Global improvement at 6 months: Global improvement in both a) pain and b) physical function will be scored using a 7-point global rating of change Likert scale with response options ranging from "much worse" to "much better" when compared to baseline. Participants indicating they are "moderately better" or "much better" will be classified as improved. All other respondents will be classified as not improved.

27. Analysis methods

Co-primary outcomes:
Main comparative analyses between groups will be performed using intention-to-treat. If more than 5% of primary outcomes are missing, multiple imputation will be applied. For the primary hypothesis, differences in mean change in pain and function (baseline minus follow-up) will be compared between groups using linear regression modelling adjusted for baseline values and the stratifying variable of radiographic severity. Results will be presented as mean differences between groups with 95% confidence intervals, and p-values will also be reported.

Secondary outcomes:
Analyses between groups will be performed using intention-to-treat. For continuous outcomes, analyses will be similar to those for the primary outcomes. Improvement based on global change scores will be compared between groups using log-binomial regression, adjusting for the stratifying variable of radiographic severity, with results reported as risk ratios and risk differences. Counts and percentages of participants experiencing improvements will be reported in each treatment group. Should the log-binomial regression models fail to converge, logistic regression models adjusting for the stratifying variable will be fit, with results reported as odds ratios, risk ratios and risk differences, calculated from fitted logistic regression models. For all between-group comparisons, 95% confidence intervals for comparisons and p-values will be reported.

28. Statistical Methods – adjustment for covariates

For the primary hypothesis, differences in mean change in pain and function (baseline minus follow-up) will be compared between groups using linear regression modelling adjusted for baseline values and the stratifying variable of radiographic severity. For continuous secondary outcomes, linear regression models will also be adjusted for baseline values and the stratifying variable of radiographic severity. For the binary outcomes, regression models will be adjusted for the stratifying variable of radiographic severity.

29. Statistical Methods – sensitivity analyses

A sensitivity analysis will estimate treatment effects on the primary outcomes assuming full adherence, where full adherence is as defined in Section 19. That is, complier average causal effects will be estimated using an instrumental variables approach (where randomization is the instrument for adherence). Two-stage least squares models will be fit [15] with complier average causal effects reported with 95% confidence intervals and p-values.

30. Statistical Methods – subgroup analyses

To assess whether the effect of shoe class on the primary outcomes is moderated by any of Kellgren & Lawrence grade, Foot Posture Index score, body mass index or baseline score on the KOOS patellofemoral pain and OA subscale, appropriate interaction terms between randomised group and each of these variables will be included in regression models for the primary outcomes, and for each potential effect modifier separately. For the continuous moderators, fractional polynomials will be explored to determine whether a more complex functional form than linear is appropriate, following the approach of Royston & Sauerbrei.[16]

The rationale for the choice of *a priori* moderators is:

- Kellgren & Lawrence grade- we hypothesise that benefits of flat flexible shoes (relative to stable supportive) will be greater in people with Grade 4 compared to Grade 3, given that people with more severe radiographic OA demonstrate higher knee adduction moment parameters than people with less severe disease, [17] and thus have greater scope for improvement with flat flexible shoes.
- Foot Posture Index score- we hypothesise that benefits of flat flexible shoes (relative to stable supportive) will be greater in people with a more pronated foot posture, given that greater knee adduction moment

reductions with other biomechanical treatments (eg lateral wedges) have been observed in people with more pronated feet, [18] thus giving such people greater scope to improve symptoms with knee-load reducing interventions.

- Body mass index- we hypothesise that benefits of flat flexible shoes (relative to stable supportive) will be greater in people with a higher body mass index, given that these people have a greater mechanical load applied to their knee joint than those with a lower body mass index, and thus have greater scope for improvement with flat flexible shoes.
- KOOS patellofemoral pain and OA subscale score- we hypothesise that benefits of flat flexible shoes (relative to stable supportive) will be greater in people with lower baseline scores on the KOOS patellofemoral pain and OA subscale (ie in those with worse patellofemoral pain), given that similar shoes reduce patellofemoral joint forces [19, 20], and therefore may benefit people with patellofemoral pain.

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

Baseline characteristics of participants with one or both primary outcomes missing at 6 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 6 months, multiple imputation will be applied. The number of imputed datasets will be approximately equal to the proportion of participants with missing primary outcomes for the primary outcome with the most missing data. 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 [21]) will be applied. We will explore the impact of the violation of the missing-at-random assumption if the assumption was violated in both groups, or in one group only.

32. Additional Analyses

Nil

33. Harms

The number (and percentage) of patients experiencing any adverse events will be presented for each treatment group and the nature of the event(s) described. An adverse event is defined as any problem experienced in the study knee or elsewhere in the body as a result of wearing the study shoes.

34. Statistical Software

Stata v15 will be used (StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC)

35. References

1. Arthritis & Osteoporosis Victoria: A Problem Worth Solving. The rising cost of musculoskeletal conditions in Australia. *Arthritis & Osteoporosis Victoria* 2013.
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, Bridgett L, Williams S, Guillemin F, Hill CL *et al*: The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 2014, 73(7):1323-1330.
3. Britt H, Miller G, Charles J, Henderson J, Bayram C, Pan Y, Valenti L, Harrison C, O'Halloran J, Fahridin S: General practice activity in Australia 2009–10. General practice series no. 27. Cat. no. GEP 27. In. Canberra: Australian Instiute of Health & Welfare; 2010.
4. National Clinical Guideline Centre: Osteoarthritis. Care and management in adults. Clinical guideline CG177. Methods, evidence and recommendations. In. London; National Institute for Health and Care Excellence; 2014.
5. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, Hawker GA, Henrotin Y, Hunter DJ, Kawaguchi H *et al*: OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis Cartilage* 2014, 22(3):363-388.
6. Andriacchi TP, Mundermann A: The role of ambulatory mechanics in the initiation and progression of knee osteoarthritis. *Current Opinion in Rheumatology* 2006, 18:514-518.
7. Chehab EF, Favre J, Erhart-Hledik JC, Andriacchi TP: Baseline knee adduction and flexion moments during walking are both associated with 5 year cartilage changes in patients with medial knee osteoarthritis. *Osteoarthritis Cartilage* 2014, 22(11):1833-1839.
8. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, Doherty M, Geenen R, Hammond A, Kjeken I *et al*: EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases* 2013, 72(7):1125-1135.
9. Bellamy N, Carette S, Ford P, Kean W, le Riche N, Lussier A, Wells G, Campbell J: Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials- results of a consensus development (Delphi) exercise. *Journal of Rheumatology* 1992, 19(3):451-457.
10. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D *et al*: Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Annals of the Rheumatic Diseases* 2005, 64(1):29-33.
11. Hinman RS, Wrigley TV, Metcalf BR, Hunter DJ, Campbell P, Paterson K, Staples MP, Bennell KL: Unloading shoes for osteoarthritis of the knee: protocol for the SHARK randomised controlled trial. *BMC Musculoskelet Disord* 2014, 15:48.
12. Schulz KF, Grimes DA: Multiplicity in randomised trials I: endpoints and treatments. *Lancet* 2005, 365(9470):1591-1595.
13. 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.
14. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010, 340:c869.
15. Stuart EA, Perry DF, Le HN, Ialongo NS: Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science* 2008, 9(4):288-298.
16. Royston P, & Sauerbrei, W.: Two Techniques for Investigating Interactions between Treatment and Continuous Covariates in Clinical Trials. *The Stata Journal* 2009, 9(2):230-251.
17. Foroughi N, Smith R, Vanwanseele B: The association of external knee adduction moment with biomechanical variables in osteoarthritis: a systematic review. *Knee* 2009, 16(5):303-309.
18. Sawada T, Kito N, Yukimune M, Tokuda K, Tanimoto K, Anan M, Takahashi M, Shinkoda K: Biomechanical effects of lateral and medial wedge insoles on unilateral weight bearing. *J Phys Ther Sci* 2016, 28(1):280-285.
19. Sinclair J: Effects of barefoot and barefoot inspired footwear on knee and ankle loading during running. *Clin Biomech (Bristol, Avon)* 2014, 29(4):395-399.
20. Ho KY, Blanchette MG, Powers CM: The influence of heel height on patellofemoral joint kinetics during

walking. *Gait Posture* 2012, 36(2):271-275.
21. 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.