Bone & Joint Open

Supplementary Material

AFTER
Ankle Fracture Treatment: Enhancing Rehabilitation

File 1. Health Economic Analysis plan.

Table of Contents

1 Background and rationale

1.1 Purpose of the health economic analysis plan

The purpose of this health economic analysis plan (HEAP) is to describe the analysis and reporting procedure intended for the economic analyses to be undertaken. The analysis plan is designed to ensure that there is no conflict with the protocol and associated statistical analysis plan (SAP) and it should be read in conjunction with them. This HEAP largely follows the guidance as suggested by Thorn et al[1]. This document has been written based on information contained in the trial protocol version 2.0, dated 22 May 2023 and SAP version 1.0, dated 14 May 2024.

1.2 Background

Ankle fracture is one of the most common fractures of the lower limb[2]. Based on a 1988-2012 analysis of UK primary care records[3], the annual incidence of ankle fracture amongst adults aged 50 years and above (104 per 100,000 person-years) was estimated to be greater than amongst adults aged 49 years and below (75 per 100,000 person-years).

Treatments for ankle fractures range from conservative plaster casts or boots to surgical fixation. Physiotherapy after surgery or a period of immobilisation in a cast or boot aims to reduce pain, enhance mobility, and improve balance and coordination as well as muscle strength. Many adults aged 50 years and above experience loss of confidence and fear of future injuries after ankle fracture[4,5]. There is good evidence that progressive exercise reduces risk of falls and improves mobility in other clinical populations[6], so a tailored progressive exercise intervention supervised by a physiotherapist could have the potential to improve recovery in this group, especially where rehabilitation needs are often more complex. However, there is insufficient evidence regarding physiotherapist-supervised rehabilitation after ankle fracture[7,8], especially, for adults aged 50 years and above.

1.3 Aims and objectives

The aim of this trial is to evaluate the clinical and cost-effectiveness of physiotherapist-supervised versus self-directed rehabilitation in improving ankle function for people aged 50 years and over with ankle fractures.

The primary objective of this trial is to quantify and draw inferences on differences in ankle function at 6 months post-randomization between the trial intervention groups (supervised vs self-directed rehabilitation).

The secondary objectives of this trial include:

- ankle function at 2 and 4 months
- ankle pain at 2, 4 and 6 months
- health-related quality of life (HRQoL) at 2, 4 and 6 months
- physical function at 4 and 6 months
- self-efficacy to exercise at 4 and 6 months
- exercise adherence at 4 and 6 months
- risk of related complications over the initial 6 months
- cost-effectiveness of the interventions

2 Trial design

AFTER is a multi-centre, parallel-group, superiority individually randomised controlled trial assessing the clinical effectiveness of supervised versus self-directed rehabilitation in improving ankle function for people aged 50 years and over after ankle fracture. The trial will be conducted at secondary care trauma departments in a minimum of 20 NHS hospitals and their related physiotherapy services.

Supervised and self-directed rehabilitation are defined as follows:

Self-directed rehabilitation. Self-directed rehabilitation is the provision of a standardised high-quality detailed advice on self-management and a set of exercises that can be progressed independently by the participant. The advice materials will be provided by a surgeon, nurse or physiotherapist during the fracture clinic appointment and will be accessible in both paper and online format. The online format comes with additional instruction videos.

Supervised rehabilitation. Participants referred to a physiotherapist will have four to six one-to-one sessions with a trial-trained physiotherapist, spread over three months from the initial session. This period allows sufficient time for neuromuscular adaptation to exercise[9]. The first session will take place as soon as possible after the referral, and no later than three weeks from randomization. These sessions will be delivered via face-to-face or telephone/videoconference, depending on the usual mode of physiotherapy delivery for the participant. Physiotherapists will support participants with a highly structure but individualised progressive exercise programme focusing on recovery of movement, muscle strength, balance and gait training, and issue the exercise programme workbook. The programme uses contemporary evidence-based guidelines on exercise volume and load to optimise the physiological response[10].

For both forms of rehabilitation, commonly used simple methods[11] to support exercise adherence such as goal setting and provision of an exercise diary will be used.

2.1 Inclusion criteria

Adults aged 50 years and over with an ankle fracture undergoing surgical fixation or non-surgical management and are provided with a cast or orthotic boot (non-removable or removable for non-weight bearing ankle movement) for at least four weeks were included.

2.2 Exclusion criteria

Participants will be excluded if they are deemed unable to adhere to trial procedures or complete questionnaires; were not ambulatory before the injury; or had contraindications to participation in an exercise programme.

3 Economic approach

3.1 Aims and objectives

The primary objective of the health economic evaluation is to estimate the within-trial cost-effectiveness of supervised versus self-directed rehabilitation among patients with ankle fracture.

3.2 Overview of economic analysis

A within-trial economic analysis will be performed using individual participant-level data from AFTER. The analytical approaches will take the form of cost-utility analysis. Based on trial evidence, incremental cost-effectiveness ratios

(ICERs) will be calculated by taking a ratio of the difference in the mean costs and mean quality-adjusted life years (QALYs). Participants will primarily be recruited from secondary care trauma departments in a minimum of 20 NHS hospitals. The economic analysis will be from the NHS and personal social services (PSS) perspective and will compare the costs and consequences of each treatment group over six months after randomization.

4 Data collection and management

4.1 Measurement of resource use

The therapist hours per treatment group will be computed from the frequency and duration of supervised and self-directed rehabilitation sessions that will be captured via the Rehabilitation Initial Provision Log and the Treatment Log for supervised rehabilitation.

Inpatient care will be recorded in the form of treatment due to complications, which will be captured via the participant-reported Complications case report form (CRF) and verified by sites.

Resource use data will be collected using trial questionnaires given to participants at two- and six-months post-randomization. Participants will be required to note the frequency of their use of outpatient care, emergency department care, community health care, private care, social care, residential/nursing home care, informal care, and painkillers related to their ankle fracture or its treatment. Participants will also have their use of walking aids and time off work recorded.

4.2 Valuation of resource use

All resource use will be valued in monetary terms using the latest and most appropriate UK unit costs or participant valuations estimated at the time of analysis. Adjustments will be made for inflation using the latest NHS hospital & community health services pay & prices index from the Personal Social Services Research Unit (PSSRU) Unit Costs compendium where applicable.

NHS reference costs will be employed to value hospital resource use (e.g. inpatient visits, emergency department visits and outpatient attendances) while the latest PSSRU Unit Costs of Health and Social Care will be used to value the therapist hours based on their grade per treatment group and community health and social services resource use. Unit cost of medications will be obtained from the Prescription Cost Analysis. The gender-specific median wage, obtained from the Office for National Statistics, will be used in the computation of the valuation of time taken off work by the participants due to the ankle fracture. The unit cost of the exercise diary and exercise programme workbook will be obtained from the trial team at its production cost.

4.3 Measurement of health outcomes

The primary health outcome measure will be QALYs derived from utility scores, obtained using the EQ-5D-5L HRQoL instrument[12]. The EQ-5D-5L instrument facilitates the generation of a utility score from a person's health related quality of life. A utility score refers to the preference that individuals have for any particular set of health outcomes. The EQ-5D-5L's descriptive system consists of five health dimensions (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression) each with five levels of severity to choose from (no problems, slight problems, moderate problems, severe problems, and extreme problems). Although not used to derive utilities, the

EQ-5D-5L HRQoL instrument also contains a visual analogue scale that records the participant's self-rated health from the worst to best health that they can imagine having. Measurement of HRQoL will be recorded using the trial questionnaires at baseline and at two-, four- and six-months post randomization.

4.4 Valuation of health outcomes

Utility scores will be derived from responses to the EQ-5D-5L descriptive system. UK utility values will be derived using the approach recommended by NICE[13], which is currently recommending the validated mapping function onto the existing EQ-5D-3L A1 tariff set developed by the Decision Support Unit[14]. QALYs will be calculated as the area under the baseline-adjusted utility curve of EQ-5D-3L utility scores using the trapezoidal rule[15].

4.5 Data cleaning

Face validity tests will be conducted on data (e.g. to identify misspelt text) and checked against the source documents. Records of resource use across different time-points will also be cross-checked to ensure that there is no duplication. Corrections made will be documented in the statistical code.

4.6 Missing data

Before carrying out the within trial analysis, the trial data will be checked for any missing data. Where possible, the reasons for missing data will be ascertained, using the methodology described by Faria et al[16], and reported. The nature and pattern of 'missingness' will be carefully considered; in particular, whether data can be treated as missing at random (MAR) and multiple imputation methods may be used if the data is MAR.

5 Data analysis

The data analysis will be conducted once all participants have been followed for six months after randomization; no interim analysis will be undertaken. The data analysis will include a within-trial analysis using imputed data that will include all randomised participants, which is in accordance with the "intention to treat" (ITT) principle, taking a six-months' time horizon.

There will be no discounting of costs and benefits as the follow-up period is less than one year. Utilisation of resource use items will be summarised by trial allocation group and follow-up period and differences between groups will be analysed using t-tests for continuous variables and Pearson chi-squared (χ^2) test for categorical variables. Means and standard errors for values of each cost category will be estimated by treatment allocation and follow-up period. Mean differences in costs and QALYs between the treatment groups will be estimated using t-tests and the bootstrap 95% confidence interval that will be computed based on 10,000 replications. This bootstrap will use Monte Carlo simulations to resample many datasets based on the original data.

Cost and QALY data will be combined to calculate incremental cost-effectiveness ratios (ICERs) and net monetary benefit (NMB) statistics from the NHS and PSS perspective in the primary economic analysis. The primary economic analysis will use incremental cost-effectiveness thresholds of £20,000 and £30,000 per QALY as recommended by NICE[17]. An additional £15,000 per QALY cost-effectiveness threshold will also be included to reflect recent trends in healthcare decision-making[18].

All analyses will be carried out using appropriate analytical software such as R or STATA. The relevant package and version number will be recorded in the health economics report.

5.1 Uncertainty analysis

A nonparametric bootstrapping approach will be used to determine the level of sampling uncertainty surrounding the mean ICER by generating 10,000 estimates of incremental costs and benefits. Decision uncertainty characterised by estimating the probability that an option is cost-effective at different cost-effectiveness thresholds will be explored using cost-effectiveness acceptability curves (CEACs).

5.2 Sensitivity analysis

Several sensitivity analyses will be undertaken to explore uncertainties surrounding key parameters in the economic evaluation as described in the following. First, the study perspective will be extended to a societal perspective which includes indirect costs (e.g. valuation of time off work). Second, complete case analysis will be used to assess the impact of missing data on the cost-effectiveness outcomes.

5.3 Subgroup analysis

As per the subgroup analysis stated in the Statistical Analysis Plan, subgroup analysis will be conducted on initial fracture management (surgical vs non-surgical).

6 Reporting of results

Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 guidelines[19] will be followed when reporting the health economic evaluation, in a format appropriate to stakeholders and policy makers. Any deviation from the HEAP will be described and justified in the final published report.

7 References

1. **Thorn J, Davies C, Brookes S, Noble S, Dritsaki M, Gray E, et al.** Content of Health Economics Analysis Plans (HEAPs) for trial-based economic evaluations: expert Delphi consensus survey. *Value in Health* 2021;24(4):539–47.

2. **Lash NJ, Fielden J, Icholas N, Ash L, Eoffrey H Orne G, Ann F Ielden And J, et al.** Ankle fractures: functional and lifestyle outcomes at 2 years. *Wiley Online Library* [Internet] 2002 [cited 24 May 2023];72(10):724–730.

3. **Curtis EM, Velde R van der, Moon RJ, Bergh JPW van den, Geusens P, Vries F de, et al.** Epidemiology of fractures in the United Kingdom 1988–2012: Variation with age, sex, geography, ethnicity and socioeconomic status. *Bone* Elsevier, 2016;87:19–26.

4. **Keene DJ, Mistry D, Nam J, Tutton E, Handley R, Morgan L, et al.** The Ankle Injury Management (AIM) trial: a pragmatic, multicentre, equivalence randomised controlled trial and economic evaluation comparing close contact casting with open surgical reduction and internal fixation in the treatment of unstable ankle fractures in patients aged over 60 years. *Health Technol Assess (Rockv)* [Internet] NIHR Journals Library, 2016 [cited 25 May 2023];20(75):1–158.

5. **Willett K, Keene DJ, Mistry D, Nam J, Tutton E, Handley R, et al.** Close Contact Casting vs Surgery for Initial Treatment of Unstable Ankle Fractures in Older Adults: A Randomized Clinical Trial. *JAMA* [Internet] American Medical Association, 2016 [cited 25 May 2023];316(14):1455–1463.

6. **Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al.** Exercise for preventing falls in older people living in the community. *Cochrane Database Syst Rev* [Internet] John Wiley and Sons, Inc. and the Cochrane Library, 2019 [cited 25 May 2023];2019(1):CD012424.

7. **Lin C-WC, Donkers NA, Refshauge KM, Beckenkamp PR, Khera K, Moseley AM**. Rehabilitation for ankle fractures in adults. *Cochrane Database Syst Rev* [Internet] Cochrane Database Syst Rev, 2012 [cited 25 May 2023];11.

8. **Moseley AM, Beckenkamp PR, Haas M, Herbert RD, Lin CWC, Evans P, et al.** Rehabilitation After Immobilization for Ankle Fracture: The EXACT Randomized Clinical Trial. *JAMA* [Internet] American Medical Association, 2015 [cited 25 May 2023];314(13):1376–1385.

9. **Byrne C, Faure C, Keene DJ, Lamb SE**. Ageing, Muscle Power and Physical Function: A Systematic Review and Implications for Pragmatic Training Interventions. *Sports Med* [Internet] Sports Med, 2016 [cited 25 May 2023];46(9):1311–1332.

10. American College of Sports Medicine position stand. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* [Internet] Med Sci Sports Exerc, 2009 [cited 25 May 2023];41(3):687–708.

11. **Abraham C, Michie S**. A taxonomy of behavior change techniques used in interventions. *Health Psychol* [Internet] Health Psychol, 2008 [cited 25 May 2023];27(3):379–387.

12. **Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, 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.

13. **National Institute for Health and Care Excellence**. NICE health technology evaluations: the manual. https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation , (date last accessed 23 February 2022).

14. **Hernandez Alava M, Pudney S, Wailoo A**. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. 2020.

15. **Drummond MF, Sculpher MJ, Torrance GW, O'Brien, Stoddart BJ and GL**. *Methods for the economic evaluation of health care programmes*. Oxford University Press, 2005.

16. **Faria R, Gomes M, Epstein D, White IR**. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *Pharmacoeconomics* 2014;32(12):1157–1170.

17. **National Institute for Health and Care Excellence**. Judging whether public health interventions offer value for money. https://www.nice.org.uk/advice/lgb10/chapter/judging-the-cost-effectiveness-of-public-health-activities , (date last accessed 1 July 2018).

18. **Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al.** Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess (Rockv)* 2015;19(14):1–504.

19. **Husereau D, Drummond M, Augustovski F, Bekker-Grob E de, Briggs A, Carswell C, et al.** Consolidated health economic evaluation reporting standards (CHEERS) 2022 explanation and elaboration: a report of the ISPOR CHEERS II Good Practices Task Force. *Value in Health* 2022;25(1):10–31.

8 Appendix A: Amendment history

Lists details of all HEAP amendments whenever a new version of the HEAP is produced.

HEAP version no., Issue date	Protocol version no., Issue date	Name of person who made the change(s)	Date changed	Significant change(s) from previous version together with reasons

File 1. Statistical analysis plan.

Contents

1 Introduction

This document details the proposed data presentation and analysis for the main paper(s) and final study reports from the ***NIHR-funded multicentre randomised controlled trial assessing the clinical effectiveness of supervised versus self-directed rehabilitation in improving ankle function for people aged 50 years and over after ankle fracture (AFTER)***. The results reported in these papers should follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial. This document follows published guidelines regarding the content of statistical analysis plans for clinical trial(1).

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this analysis strategy. If reported, the analyses will be marked as post-hoc; the source of the suggestion will be acknowledged and the reader will be advised to rely on the pre-specified analysis for the interpretation of the results.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified, and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

Integral to this Statistical Analysis Plan (SAP) is the *SAP – Data Definitions and Tables* document which will include full detailed descriptions of all key outcomes, including their definition, generation and how they will be reported at the end of the study. These two documents should be read in tandem.

1.1 Key personnel

Author(s) (Trial statistician)	Dr Asma Saleh Email: asma.saleh@ndorms.ox.ac.uk Tianshu Liu (previous statistician)
Author (Trial health economist)	May Ee Png Email: may.png@phc.ox.ac.uk
Reviewer (Trial Coordinator)	Kylea Draper

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS)
Kadoorie Centre,
Level 3, John Radcliffe Hospital
Oxford, OX3 9DU
Email: kylea.draper@ndorms.ox.ac.uk

Reviewer (TOC Members)

Andrew Duckworth
Email: Andrew.Duckworth@ed.ac.uk

Farzana Kausir
Email: kausirfarzana@gmail.com

Ines Rombach
Email: i.rombach@sheffield.ac.uk

Jonathan Hill
Email: j.hill@keele.ac.uk

Matthew Costa (AFTER trial Co-Lead)
Email: matthew.costa@ndorms.ox.ac.uk

Approver (Senior Statistician)

Nicholas Peckham
Email: nicholas.peckham@ndorms.ox.ac.uk

Approver (Chief Investigator)

David Keene
Nuffield Dept of Orthopaedics Rheumatology and Musculoskeletal Sciences
Nuffield Orthopaedic Centre
Windmill Road
OXFORD OX3 7LD

Email: david.keene@ndorms.ox.ac.uk

1.2 Changes from previous version of SAP

A summary of key changes from earlier versions of SAP, with particular relevance to protocol changes that have an impact on the design, definition, sample size, data quality/collection and analysis of the outcomes will be provided. Include protocol version number and date.

Version number Issue date	Author of this issue	Protocol Version & Issue date	Significant changes from previous version together with reasons
V1.0_14May2024		Protocol_V2.0_22May2023	Not applicable as this is the 1st issue
			Add to or delete as required

2 Background and objectives

Ankle fractures are very common and the incidence of these fractures in the UK is highest among people aged 50 years and over (2, 3). As the population ages, a three-fold increase in these fractures is projected over the next two decades (4). The mechanism of injury for people aged over 50 is usually a fall from standing height; the fracture is then defined as a fragility fracture (5).

Treatments for ankle fractures range from non-surgical management through application of plaster casts or boots to surgical fixation. A recent trial including adults aged 60 years and over found that, regardless of the initial fracture management, post-injury reduced ankle function and walking abnormalities remain at 6 months post-injury (6, 7). Ankle movement restrictions during weight bearing are usually lifted by the orthopaedic team six weeks after injury. At this stage, national guidance is that patients should be advised to gradually resume activities. Data has shown that, in addition to the advice, some patients are referred to see a physiotherapist for supervised rehabilitation. Physiotherapy after ankle fracture aims to support patients during the recovery period and provide individualised progressive exercise to improve muscle strength, range of motion, gait and balance. However, the referral patterns vary from hospital to hospital (7) and there is no national guidance on whether further rehabilitation under the supervision of a physiotherapist should be provided.

Previous research in younger adults found additional physiotherapy did not improve recovery, but it is not clear whether older adults would benefit. Moreover, extra physiotherapy could bring cost burden to the health service.

Therefore, the aim of this pragmatic, parallel-group, randomised controlled superiority trial is to evaluate the clinical and cost-effectiveness of supervised versus self-directed rehabilitation in improving ankle function for people aged 50 years and over with ankle fractures.

2.1 Objectives

Table 1 Primary and secondary objectives and endpoints.

	Objectives	Outcome Measures	Timepoint(s)
Primary	Ankle Function	Olerud and Molander Ankle Scale (OMAS)	6 months post-randomization
Secondary	Ankle Function	OMAS	Baseline, 2 and 4 months post-randomization
	Health-related quality of life	EQ-5D-5L	Baseline, 2, 4 and 6 months post-randomization
	Pain	Pain sub-scales of the EQ-5D-5L and OMAS	Baseline, 2, 4 and 6 months post-randomization
	Physical Function	PROMIS Physical Function	Baseline, 4 and 6 months post-randomization
	Self-efficacy	Self-Efficacy Exercise Score (SEE)	Baseline, 4 and 6 months post-randomization
	Exercise adherence	Self-reported exercise frequency	2, 4 and 6 months post-randomization
	Complications	Complications Questionnaire and Case Report Form	2, 4 and 6 months post-randomization
	Resource use	Health economics questionnaire	2 and 6 months post-randomization

3 Study METHODS

3.1 Trial Design/framework

AFTER is a multi-centre, parallel-group, superiority individually randomised controlled trial assessing the clinical effectiveness of supervised versus self-directed rehabilitation in improving ankle function for people aged 50 years and over after an ankle fracture. The trial will be conducted at secondary care trauma departments in a minimum of 20 NHS hospitals and their related physiotherapy services.

3.2 Randomization and Blinding

The randomization will be on a 1:1 basis to supervised versus self-directed rehabilitation, using a validated computer randomization program managed through a secure (encrypted) web-based service by the Oxford Clinical Trials Research Unit (OCTRU). Randomization will use a minimisation algorithm to ensure balanced allocation across the two treatment groups, stratified by centre and initial fracture management (surgical vs non-surgical). The first few participants will be randomised by simple randomization to seed the minimisation algorithm and a probabilistic element introduced to the algorithm to ensure the unpredictability of intervention allocation.

Full details of the randomization are available in AFTER_RandomizationAndBlindingPlan_V1.0_27Jul2022, stored in the confidential statistical section of the Trial Master File (TMF).

Due to the nature of the intervention, it is not possible to blind participants or those delivering the intervention. The local research team at recruiting centres will also not be blinded to the treatment allocation.

3.3 Sample Size

292 (146 per arm) participants providing primary outcome data at 6 months are required to detect a difference of 8 points on the OMAS score with an estimated standard deviation of 21 with 90% power and 5% (2-sided) significance. The minimum clinically important difference for the OMAS selected in surgical trials has usually been 10 points but for this trial of physiotherapy we have chosen a smaller difference of 8 points, which is likely to be clinically important and was supported by our patient advisory group. The chosen standard deviation of 21 is based on the AIM trial (6) (SD 21.7) and the feasibility study data (SD 20.5 based on 32 participants having reached the 6 month time-point). This equates to a standardised effect size of 0.38, a small to moderate effect. In the AFTER feasibility study there was 11% loss to follow-up (those not providing the primary outcome data). In order to allow for potential loss to follow-up of participants in the definitive trial we have inflated the sample size by 15% to aim for a minimum of 344 participants (172 per arm).

3.4 Statistical Interim Analysis, Data Review and Stopping guidelines

The study conduct and data safety will be discussed in the Trial Oversight Committee (TOC). The Trial Oversight Committee (TOC), which includes independent members, provides overall supervision of the trial on behalf of the funder. Its terms of reference is drawn up in a TOC charter (stored in AFTER_TOCCharter_V1.0_18Aug2022) which outlines its roles and responsibilities. Full details of the data planned to be reported at the TOC meeting are available in the TOC report template (AFTER_TOCTemplate_V1.0_17Aug2022).

There is no interim analysis planned for this trial. The trial is considered low risk so does not have a separate Data Safety Monitoring Committee (DSMC).

3.5

3.6 Timing of Final Analysis

The final analysis will be conducted after the last participant has reached the 6 month time period, and the data has been received, cleaned, and finalised.

3.7 Blinded analysis

A blinded analysis (i.e. a review of the data prior to details of the interventions being added to the dataset) will be performed to look into the distribution of variables, missing data distributions, outliers and to finalise the per protocol population. The blinded analysis will be undertaken prior to the final datalock.

3.8 Statistical Analysis Outline

The primary analysis population will be the intention-to-treat population (ITT), that is participants will be analysed in the group in which they were randomised regardless of what treatment they received. The analyses will be repeated for the per protocol (PP) population. The per protocol population will exclude participants who did not receive the allocated treatment and participants with major protocol deviations, this will be finalised using blinded data prior to

the final datalock. Standard descriptive statistics will be used to describe the demographics between the treatment groups reporting means and standard deviations or medians and interquartile ranges as appropriate for continuous variables and numbers and percentages for binary and categorical variables. Standard statistical summaries and graphical plots will be presented for the primary outcome measure and all secondary outcome measures.

The OMAS score at 6 months is the primary outcome in this study and will be compared between treatment groups as the dependent variable in a mixed-effects linear regression model including outcome information at intermediate time-points. This model will adjust for stratification factors (initial fracture management; operative or non-operative) and baseline OMAS score. A random effect will be included to account for heterogeneity due to recruitment centres. The treatment effect will be based on adjusted mean differences at 6 months and will be reported together with their 95% confidence intervals.

We will also undertake a Complier Average Causal Effect (CACE) analysis which essentially compares the "compliers" in each group. The CACE analysis will be undertaken using the definition of full compliance in the intervention group – defined as receiving a minimum of 4 physiotherapy sessions. Partial compliance is receiving at least one physiotherapy session (i.e. starting the treatment), which will be used in the per-protocol analysis. This will provide supporting evidence to any findings from the principal analysis. Subgroup analysis by surgical vs non-surgical treatment of the fracture will be undertaken, using a treatment by subgroup interaction term and will be presented using forest plots.

Similar methods to the primary outcome will be used to analyse continuous secondary clinical outcomes and patient reported outcomes. Complications will be reported by type for each intervention group, and, if appropriate, compared between the two groups using logistic regression models.

4 Statistical Principles

4.1 Statistical Significance and Multiple Testing

There is no multiple testing as only a single primary outcome is considered. Therefore, the significance levels used will be at the 2-sided 5% level and 95% confidence intervals will be reported. P-values will not be reported for subgroup exploratory analyses, the analyses results will be presented as forest plots with 95% confidence intervals. Any analyses that are not pre-specified will use the stricter 1% significance level and will be stated clearly as not included in the SAP.

4.2 Definition of Analysis Populations

The primary analysis population will be the intention-to-treat population (ITT), that is participants will be analysed in the randomised group in which they were allocated to regardless of what treatment they received. Patients with baseline outcome and at least one post-randomization outcome will be included in the analysis.

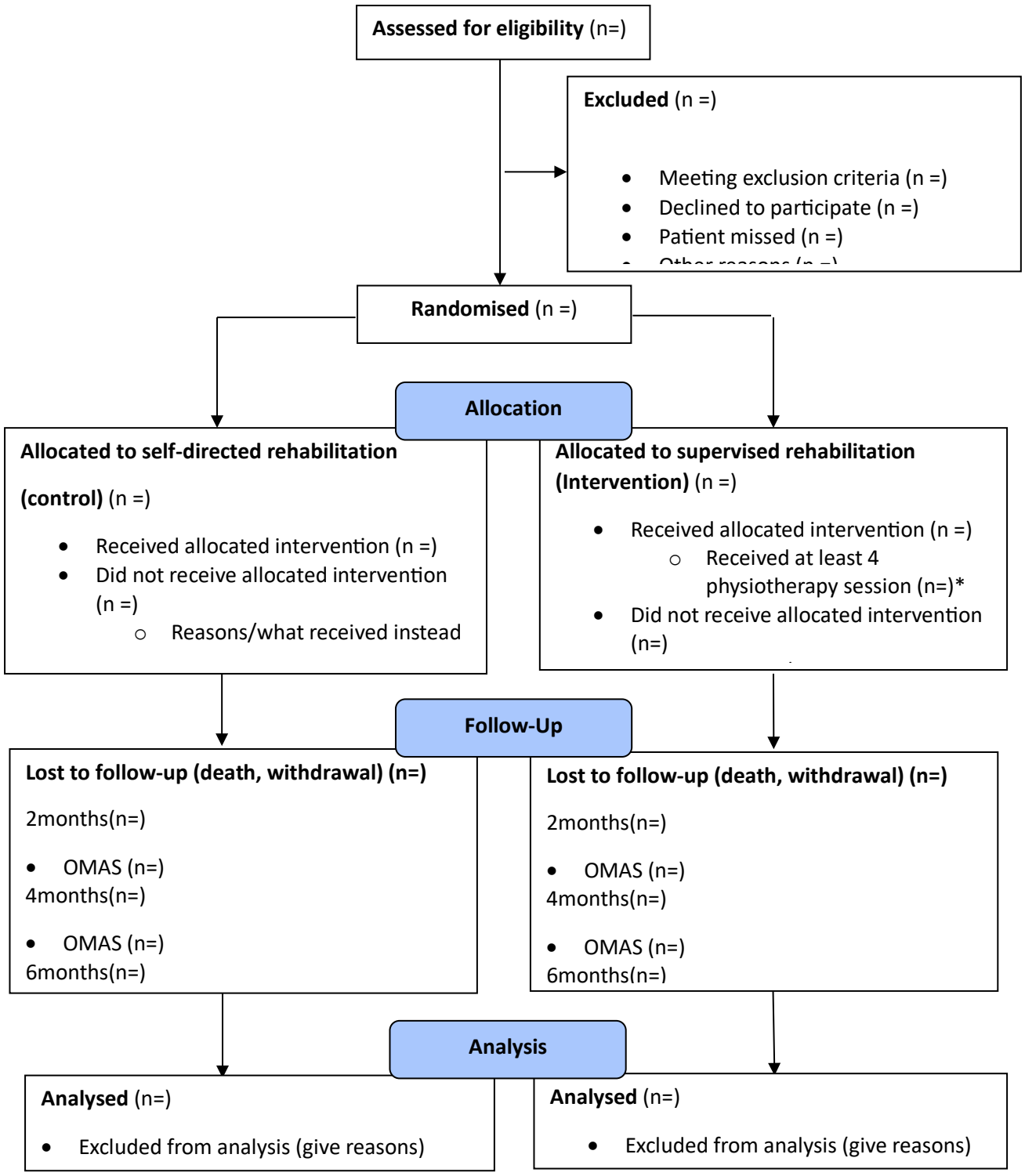
The analyses will be repeated for the per protocol (PP) population. The per protocol population will exclude participants who did not receive the allocated treatment and participants with major protocol deviations.

5 trial population and descriptive analyses

Summary of flow of trial participants through the trial and baseline stratification, demographic and clinical characteristics of each group.

5.1 Representativeness of Study Sample and Patient Throughput

Figure 1 shows the flow of participants through each stage of the trial, including numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome, following the appropriate guideline (e.g. CONSORT). Information on the number of patients screened and found to be ineligible because they meet the exclusion criteria will also be reported.



* *Full adherence in the intervention group is defined as receiving a minimum of 4 physiotherapy sessions and partial adherence is receiving at least one physiotherapy session (i.e. starting the treatment)*

Figure 1 The CONSORT flowchart.

5.2 Withdrawals

The number (and percentage) of withdrawal reasons will be presented by treatment arm. Details of each withdrawal will also be investigated, including their days on the trial, site, last questionnaire completed and whether they have their primary outcome available for analyses. (see table 7 and 8 in AFTER_SAP_DDT_V1.0_18Aug2022). Any deaths (and their causes) will be reported separately (see table 12 in the AFTER_SAP_DDT_V1.0_18Aug2022).

5.3 Baseline Characteristics

Baseline comparability between two treatment groups (supervised rehabilitation and self-directed rehabilitation) will be investigated. Numbers (with percentages) for binary and categorical variables and mean (and standard deviation),

or median (with lower and upper quartiles) for continuous variables will be presented; there will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any baseline variable (see Table 5 in AFTER_SAP_DDT_V1.0_18Aug2022) .

5.4 Unblinding

AFTER is not a blinded trial.

5.5 Treatment Compliance with Details of Intervention

A summary of treatment received (supervised rehabilitation and self-directed rehabilitation) will be provided by treatment arm. Compliance to the supervised rehabilitation treatment is defined as receiving a minimum of 4 physiotherapy sessions and partial compliance is receiving at least one physiotherapy session (i.e. starting the treatment). Compliance will be summarised as full compliance and partial compliance by treatment arm. The reasons for participants who did not receive the treatment that they were allocated will be also summarised by treatment arm (see table 6 in AFTER_SAP_DDT_V1.0_18Aug2022). Full details of intervention delivery will be reported in line with the TIDieR checklist for reporting complex interventions (8).

5.6 Reliability

To ensure consistency, validation checks of the data will be conducted. This will include checking for duplicate records, checking the range of variable values and validating potential outliers where possible (referring back to sites if necessary). As the data is collected electronically, many of these checks will be implemented automatically as part of the data entry procedure and data collection instruments have been validated prior to data entry commencing. Calculations and processes performed by a computer program, including the construction of derived data, will be checked. A sample of 15 participants will be randomly selected from all randomised participants (around 5% of the total planned sample size) for the check. Algorithms and calculations that are performed by a computer program will be checked and validated manually for the 15 selected participants.

For each variable, missing value codes will be checked for consistency and the proportion of missing values per variable will be presented. Patterns of missing data will be explored.

6 Analysis

6.1 Outcome Definitions

6.1.1 Primary outcome

The primary outcome is patient-reported ankle-related symptoms and function at 6 months after randomization measured by completion of the Olerud and Molander Ankle Score (OMAS) (9). The OMAS is a 9-item questionnaire which is completed directly by the participant (0-100, with higher scores indicating better function). The OMAS has been the primary outcome for a number of other ankle fracture trials, including NIHR HTA trials. (7, 10)

6.1.2 Secondary outcomes

Ankle Function: measured by completion of the Olerud and Molander Ankle Score (OMAS) at baseline, 2 and 4 months after randomization.

Health-related quality of life: assessed using the EuroQol 5 Dimensions (EQ-5D-5L) score (11). The EQ-5D-5L is a validated, generalised and standardised instrument comprising a VAS measuring self-rated health and a health status instrument, consisting of a five-level response (no problems, some problems, moderate problems, severe problems and unable) for five domains related to daily activities; mobility, self-care, usual activities, pain and discomfort, and anxiety and depression. Responses to the health status classification system are converted into an overall score using a published utility algorithm for the UK population. The EQ-5D health status scale ranges from negative scores reflecting a patient's quality of life being worse than death, and 0 [death] to 1 [perfect health]. A respondent's EQ-VAS gives self-rated health on a scale where the endpoints are labelled 'best imaginable health state' (100) and 'worst imaginable health state' (0). This is measured at baseline, 2, 4 and 6 months after randomization.

Pain: assessed using the sub-scales of the OMAS and EQ-5D-5L. This is measured at baseline, 2, 4 and 6 months after randomization.

Physical function: assessed using PROMIS Physical Function (12). Patient Reported Outcome Measurement Information System (PROMIS) questionnaires are administered electronically, via a Computer Adaptive Test (CAT),

which are dynamic tests based on Item Response Theory (IRT). A mathematical model adapts the sequential questions asked based on a participant's previous response. A tailored set of questions is therefore asked from a large item pool. Participants are typically asked 4 to 6 questions. PROMIS instruments are scored from 0 to 100 with 50 points representing the mean score for the US general population, higher scores indicate better function. Participants who have not completed the online questionnaire or who have no internet access will be able to complete a paper-based version of the PROMIS questionnaire with 4-items (PROMIS Physical Function - Short Form 4a) via postal follow up. If a participant needs to be contacted directly by phone to complete their follow up, they will be asked the PROMIS Physical Function CAT questionnaire as the central site team can directly enter patient responses on their behalf. This is measured at baseline, 4 and 6 months after randomization.

Self-efficacy: assessed using the Self-Efficacy Exercise Score (13). The score measures a participant's judgment of their confidence to carry out exercise. The questionnaire has 9-items specifically about the ability to continue to exercise despite barriers. The participant scores their confidence level from 0 (not confident) to 10 (very confident), if they were to exercise 3 times per week for 20 minutes during each of the nine situations presented. The overall scores range 0 to 90, with higher scores indicating greater confidence to exercise. This is measured at baseline, 4 and 6 months after randomization.

Exercise adherence: assessed using patient reported exercise frequency. This is measured at 2, 4 and 6 months after randomization.

Complications: fracture and treatment complications will be recorded, but particular note will be made of complications related to the interventions. This is measured at 2, 4 and 6 months after randomization.

Resource use: patient reported resource use and information on hospital treatments and appointments will be collected. This will include consultations with primary and secondary care, prescribed and over-the-counter pain medication use, additional physiotherapy and hospital admission, self-funded health and social care, out-of-pocket expenses, and work absence. This is measured at 2 and 6 months after randomization.

6.2 Analysis Methods

6.2.1 Primary outcome analysis

The OMAS at 6 months is the primary outcome. Raw summary statistics (mean and standard deviation) will be calculated and reported by treatment arm for each time point.

The treatment difference between the two groups (self-directed rehabilitation and supervised rehabilitation) will be analysed using a mixed effect linear regression model. The OMAS at all follow-up time points will be the dependent variable (panel data) and the independent variable will be the interaction between treatment group and follow-up time points. This model will adjust for stratification factors of randomization (recruitment centre and initial fracture management (operative vs non-operative)), age in years, and sex at birth as fixed effects in the model.

Sensitivity analyses will be conducted to check the robustness of the analysis results. The analyses will include a partially-adjusted model, per protocol analysis and CACE analysis. A partially-adjusted model will only include baseline OMAS as its sole adjustment variable. Per-protocol analysis will only include participants at least partially compliant in the supervised rehabilitation arm to their allocated treatment (1 or more physiotherapy sessions in supervised rehabilitation arm), and will be conducted for the adjusted model. The per-protocol analysis will also exclude participants who did not receive the self-directed rehabilitation materials when allocated to self-directed rehabilitation.

Complier Average Causal Effect (CACE) analysis:

CACE analysis seeks to compare outcomes for individuals in the intervention condition who complied with treatment with individuals in the control group who would have complied with treatment given the opportunity to do so. The two treatment groups for the AFTER trial are supervised rehabilitation (intervention) and self-directed rehabilitation (control).

Participants in the intervention arm will be defined as non-compliers to the control arm if they have not met the adherence definition: Full compliance is defined as receiving a minimum of 4 physiotherapy sessions. However, it is unlikely for participants in the control arm to crossover to the intervention arm, therefore the per-protocol analysis

will mostly exclude non-compliers in the intervention arm that may result in estimation bias. CACE looks at compliers in the intervention arm and would-be compliers in the control arm (i.e. the participants in the control arm who would comply to the treatment if they were assigned to the intervention arm) (14). The CACE analysis will be done using a two-stage regression model (14-16):

Stage 1: Predict treatment compliance using treatment allocated. Treatment received (compliance) by participants will be used as an instrumental variable (IV) for treatment allocated. A mixed model, with the independent variable as treatment allocation and the dependent variable as treatment compliance, will run, adjusting for baseline OMAS and initial fracture management. The coefficient estimated for treatment allocation will be used to predict treatment compliance in both intervention and control arm.

Stage 2: Use predicted treatment compliance to represent treatment allocation. A mixed effect model will be run with the dependent variable as OMAS scores and independent variable as predicted treatment compliance and follow-up timepoint interaction. The random effect of the model will be recruitment centre and the model will adjust for fracture management and baseline OMAS score. The Stata command "ivregress" can perform the two-stage analysis together.

For fully adjusted (as per primary outcome analysis) and non-adjusted model, the process will be the same with different variables being adjusted.

The treatment effect for OMAS at 6 months (ITT) and sensitivity analyses results will be presented together in one table for comparison (see table 9 in AFTER_SAP_DDT_V1.0_18Aug2022). They will be reported alongside 95% confidence intervals and p-values calculated using clustered standard error to account for serial correlations. Mixed model assumptions will be assessed using residual plots. In case we have a skewed variable, we will consider transforming or using bootstrapping methods.

6.2.2 Secondary outcome analysis

Summary statistics by treatment arm will be computed for EQ-5D-5L, pain, PROMIS, self-efficacy, exercise adherence and complications. Similar methods to the primary outcome will be used to analyse continuous secondary outcomes, with their mean and standard deviation or median and interquartile range reported. Only the intention to treat (ITT) population will be used for secondary outcome analysis (see table 10 in AFTER_SAP_DDT_V1.0_18Aug2022).

OMAS at baseline, 2 and 4 months: The treatment difference for the OMAS at other follow-up time points will be the adjusted difference at these time points taken from the same model used for the primary outcome analysis.

EQ-5D-5L utility score (at baseline, 2,4 and 6 months): will be calculated by the ED-5D-5L mapping function (17) and range from -0.594 to 1. If a participant dies before end of follow-up, their utility score will be imputed as 0. Similar to the analysis for primary outcome, a mixed effect model with recruitment centre as random effect and adjust for stratification factor (initial fracture management) and baseline utility score. The independent variable of the model will be treatment and follow-up time point interaction. This model will adjust for stratification factors of randomization (recruitment centre and initial fracture management (operative vs non-operative), as well as important prognostic factors of age and sex. Results for treatment effect at each of the follow-up timepoint will be presented with confidence intervals and p-values.

EQ-5D-5L VAS score (at baseline, 2, 4 and 6 months): is a score ranges from 0 to 100, with better score indicating better health. The analysis for the VAS score will be the same as with the EQ-5D-5L utility score, but adjusting for baseline EQ-5D-5L VAS score.

PROMIS Physical Function (at baseline, 4 and 6 months): assesses the physical function of individuals. A normalised score will automatically be calculated by REDCap built-in programmes. Similar to the analysis for the primary outcome, a mixed effect model with recruitment centre as random effect and adjust for stratification factor (initial fracture management) and baseline PROMIS score. The independent variable of the model will be treatment and follow-up time point interaction. This model will adjust for stratification factors of randomization (recruitment centre and initial fracture management (operative vs non-operative) , as well as important prognostic factors of age and sex. Participants who have no access to the online PROMIS computer-adaptive test questionnaire can complete a paper-based PROMIS short form. A sensitivity analysis excluding the PROMIS short form will be conducted to assess

the robustness of the analysis. Results for treatment effect at each of the follow-up timepoints will be presented with 95% confidence intervals and p-values.

Pain (at baseline, 2, 4 and 6 months) will be assessed by both the OMAS and EQ-5D-5L subscale. OMAS pain has 5 categories and ranges from 0-25, the pain decreases as score increases. EQ-5D-5L pain has 5 categories and ranges from 1-5, the pain increases as the score decreases. Both scores will be analysed using a mixed ordinal regression model, adjusting for centre as a random effect and fixed effects for fracture management, their respective baseline score, as well as important prognostic factors of age and sex.

Self-efficacy for exercise (SEE) score (at baseline, 4 and 6 months): ranges from 0-90 with 90 being the most confident and 0 being the least confident for conducting exercise. A mixed effect model with recruitment centre as random effect and adjust for stratification factor (initial fracture management) and baseline SEE score, as well as important prognostic factors of age and sex. The independent variable of the model will be treatment and follow-up time point interaction.

Exercise frequency (at 2, 4 and 6 months): records the frequency of exercise that participants perform in a week, ranges from 7 days to less than 1 day a week. Summary statistics by treatment arm will be presented.

Complications: will be reported by participants and verified by sites. Both patient-reported complications (at 2, 4 and 6 months) and site-verified complications (ad-hoc) will be summarised by treatment arm (see table 11 and 13 in AFTER_SAP_DDT_V1.0_18Aug2022). The number and its associated percentage will be reported for each complication category. A logistic regression model will be used to analyse this outcome variable.

6.3 Missing Data

Missing data will be minimised by careful data management. Missing data will be described with reasons given where available; the number and percentage of individuals in the missing category will be presented by treatment arm. All data collected on data collection forms will be used, since only essential data items will be collected. No data will be considered spurious in the analysis since all data will be checked and cleaned before analysis.

The primary analysis method proposed is reasonably robust to missing at random (MAR) data (18), but this assumption will be investigated by using a pattern mixture model, such as the Stata 'rctmiss' command. The missing data will be assumed to have the same distribution as the observed data, differing by a sensitivity parameter delta. The delta parameter will be the mean shifted value between observed and missing data, measuring the departure from MAR up to one standard deviation from the observed data value. This missing data sensitivity analysis can be used to explore the impact of any potential differential differences in the amount of missing data in the allocated treatment arms.

6.4 Pre-specified Subgroup Analysis

Subgroup analysis will be conducted on fracture management (surgical vs non-surgical). The purpose of subgroup analyses is to investigate if the estimated treatment effects are relatively consistent across subgroup and for this extent will be viewed as exploratory.

The analysis population will be divided into their respective subgroup, and if there are too few participants in one subgroup (<=15) or one treatment arm of a subgroup (<=5), the analysis of that subgroup is likely to be biased due to small number of samples. In that case, that subgroup will be combined with other subgroups with more participants or the subgroup analysis will not be conducted.

Linear regression will be performed for each subgroup analysis. The model will include a treatment-subgroup interaction term to capture the effect of subgroups. The estimated treatment difference and confidence intervals will be reported and presented in forest plots. The results will be presented in forest plots with 95% confidence interval.

6.5 Supplementary/ Additional Analyses and Outcomes

The primary analysis ignores the potential clustering effect on outcomes of patients treated by the same therapist. The intra-cluster correlation coefficient (ICC) will be presented to indicate if the assumption of ignoring the therapist effect is fair for the primary analysis. We plan to test this assumption and perform a sensitivity analysis; this will be done by modelling the clustering therapist effect in the intervention arm as a random effect, and by treating the control arm as a single cluster. We will also explore, through the use of forest plots, the consistency of the therapist effect. (19, 20)

6.6 Harms

Serious adverse events (SAEs) will be reported by treatment arm and overall, in numbers and percentages (see table 14 in AFTER_SAP_DDT_V1.0_18Aug2022). Complications such as wound infection, nerve injury, pressure sores and deep vein thrombosis and embolism will be reported by participants and verified by sites.

6.7 Meta-analyses (if applicable)

There is no planned meta-analysis in this study.

7 Validation of the Primary analysis

To validate the primary outcome (OMAS) and key secondary outcomes (PROMIS Physical Function and EQ5D-5L) a statistician not involved in the trial will independently repeat the analyses detailed in this SAP, by using different statistical software (if possible). The results will be compared and any unresolved discrepancies will be reported in the Statistical report (See OCTRU SOP STATS-005 Statistical Report). If necessary this will include derivation of the primary and key secondary outcomes from raw data.

8 Specification of Statistical Packages

All analysis will be carried out using appropriate validated statistical software such as STATA, SAS, SPLUS or R. The relevant package and version number will be recorded in the Statistical report.

9 Publication

This study will be conducted as part of the portfolio of trials in the registered UKCRC Oxford Clinical Trials Research Unit (OCTRU) at the University of Oxford. It will follow their Standard Operating Procedures ensuring compliance with the principles of Good Clinical Practice and the Declaration of Helsinki and any applicable regulatory requirements.

10 Appendix: glossary of abbreviations

AE	Adverse event
CACE	Compliance Average Causal Effect
CI	Confidence Interval
CRF	Case Report Form
CT	Computerised Tomography Scan
CTRG	Clinical Trials and Research Governance
DSMC	Data Safety and Monitoring Committee
EQ-5D	EuroQoL Five Dimensions
GCP	Good Clinical Practice
GP	General Practitioner
HRA	Health Research Authority
IV	Instrumental variable
NHS	National Health Service
NIHR	National Institute for Heath and Care Research
OMAS	Olreud-Molander Ankle Scale
OR	Odds Ratio
ORTU	Oxford Respiratory Trials Unit

PI	Principal Investigator
PP	Per Protocol
PROM	Patient-Reported Outcome Measure
PROMIS	Patient-Reported Outcomes Measurement Information System
QoL	Quality of Life
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
REDCap	Research Electronic Data Capture
RFPB	Research for Patient Benefit
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SEE	Self-Efficacy for Exercise Scale
TOC	Trial Oversight Committee
TSC	Trial Steering Committee
SOP	Standard Operating Procedure
VAS	Visual Analogue Scale

Section/Item	Index	Description	Reported on page #
Section 1: Administrative information			
Trial and Trial registration	1a	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle, and trial acronym (if applicable)	1
	1b	Trial registration number	1
SAP Version	2	SAP version number with dates	1
Protocol Version	3	Reference to version of protocol being used	1
SAP revisions	4a	SAP revision history	4
	4b	Justification for each SAP revision	4
	4c	Timing of SAP revisions in relation to interim analyses, etc.	4
Roles and responsibility	5	Names, affiliations, and roles of SAP contributors	3-4

Signatures of:	6a	Person writing the SAP	Saved in Statistical TMF
	6b	Senior statistician responsible	Saved in Statistical TMF
	6c	Chief investigator/clinical lead	Saved in Statistical TMF
Section 2: Introduction			
Background and rationale	7	Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial	5
Objectives	8	Description of specific objectives or hypotheses	5-6
Section 3: Study Methods			
Trial design	9	Brief description of trial design including type of trial (e.g., parallel group, multi-arm, crossover, factorial) and allocation ratio and may include brief description of interventions	6
Randomization	10	Randomization details, e.g., whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)	6
Sample size	11	Full sample size calculation or reference to sample size calculation in protocol (instead of replication in SAP)	6
Framework	12	Superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis	6
Statistical interim analysis and stopping guidance	13a	Information on interim analyses specifying what interim analyses will be carried out and listing of time points	6-7
	13b	Any planned adjustment of the significance level due to interim analysis	6-7
	13c	Details of guidelines for stopping the trial early	6-7
Timing of final analysis	14	Timing of final analysis, e.g., all outcomes analysed collectively or timing stratified by planned length of follow-up	7

Timing of outcome assessments	15	Time points at which the outcomes are measured including visit "windows"	7
Section 4: Statistical Principals			
Confidence intervals and *P* values	16	Level of statistical significance	7-8
	17	Description and rationale for any adjustment for multiplicity and, if so, detailing how the type 1 error is to be controlled	7-8
	18	Confidence intervals to be reported	7-8
Adherence and Protocol deviations	19a	Definition of adherence to the intervention and how this is assessed including extent of exposure	14
	19b	Description of how adherence to the intervention will be presented	14
	19c	Definition of protocol deviations for the trial	11
	19d	Description of which protocol deviations will be summarized	11
Analysis populations	20	Definition of analysis populations, e.g., intention to treat, per protocol, complete case, safety	8
Section 5: Trial Population			
Screening data	21	Reporting of screening data (if collected) to describe representativeness of trial sample	10
Eligibility	22	Summary of eligibility criteria	10
Recruitment	23	Information to be included in the CONSORT flow diagram	9
Withdrawal/ Follow-up	24a	Level of withdrawal, e.g., from intervention and/or from follow-up	11
	24b	Timing of withdrawal/lost to follow-up data	11
	24c	Reasons and details of how withdrawal/lost to follow-up data will be presented	11
Baseline patient characteristics	25a	List of baseline characteristics to be summarized	12
	25b	Details of how baseline characteristics will be descriptively summarized	12
Section 6: Analysis			

Outcome definitions		List and describe each primary and secondary outcome including details of:	
	26a	Specification of outcomes and timings. If applicable include the order of importance of primary or key secondary end points (e.g., order in which they will be tested)	12-13
	26b	Specific measurement and units (e.g., glucose control, hbA1c [mmol/mol or %])	12-13
	26c	Any calculation or transformation used to derive the outcome (e.g., change from baseline, QoL score, Time to event, logarithm, etc.)	12-13
Analysis methods	27a	What analysis method will be used and how the treatment effects will be presented	13-16
	27b	Any adjustment for covariates	13-16
	27c	Methods used for assumptions to be checked for statistical methods	13-16
	27d	Details of alternative methods to be used if distributional assumptions do not hold, e.g., normality, proportional hazards, etc.	13-16
	27e	Any planned sensitivity analyses for each outcome where applicable	13-16
	27f	Any planned subgroup analyses for each outcome including how subgroups are defined	13-16
Missing data	28	Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)	16
Additional analyses	29	Details of any additional statistical analyses required, e.g., complier-average causal effect10 analysis	14,16
Harms	30	Sufficient detail on summarizing safety data, e.g., information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analysed, i.e., grade 3/4 only, incidence case analysis, intervention emergent analysis	16-17
Statistical software	31	Details of statistical packages to be used to carry out analyses	18
References	32a	References to be provided for nonstandard statistical methods	18-19
	32b	Reference to Data Management Plan	18-19
	32c	Reference to the Trial Master File and Statistical Master File	18-19
	32d	Reference to other standard operating procedures or documents to be adhered to	18-19

Taken from the paper: Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. 2017;318(23):2337-43.

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; hbA1c, haemoglobin A1c; QoL, quality of life; SAP, statistical analysis plan.

11 References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical TrialsGuidelines for the Content of Statistical Analysis Plans in Clinical TrialsGuidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-43.
2. Court-Brown CM, Caesar B. Epidemiology of adult fractures: a review. Injury. 2006;37(8):691-7.
3. Curtis EM, van der Velde R, Moon RJ, van den Bergh JP, Geusens P, de Vries F, et al. Epidemiology of fractures in the United Kingdom 1988–2012: variation with age, sex, geography, ethnicity and socioeconomic status. Bone. 2016;87:19-26.
4. Kannus P, Palvanen M, Niemi S, Parkkari J, Järvinen M. Stabilizing incidence of low-trauma ankle fractures in elderly people: Finnish statistics in 1970–2006 and prediction for the future. Bone. 2008;43(2):340-2.
5. Health NIf, Excellence C. Osteoporosis: assessing the risk of fragility fracture: National Institute for Health and Clinical Excellence; 2012.
6. Willett K, Keene DJ, Mistry D, Nam J, Tutton E, Handley R, et al. Close contact casting vs surgery for initial treatment of unstable ankle fractures in older adults: a randomized clinical trial. Jama. 2016;316(14):1455-63.
7. Willett K, Keene DJ, Morgan L, Gray B, Handley R, Chesser T, et al. Ankle Injury Management (AIM): design of a pragmatic multi-centre equivalence randomised controlled trial comparing Close Contact Casting (CCC) to Open surgical Reduction and Internal Fixation (ORIF) in the treatment of unstable ankle fractures in patients over 60 years. BMC musculoskeletal disorders. 2014;15:1-9.
8. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ : British Medical Journal. 2014;348(mar07 3).
9. Olerud C, Molander H. A scoring scale for symptom evaluation after ankle fracture. Archives of orthopaedic and traumatic surgery. 1984;103(3):190-4.
10. Griffin X, Kearney R, Keene DJ, Marques E, Achten J, Pinedo-Villanueva R, et al. In younger adults with unstable ankle fractures treated with close contact casting, is ankle function not worse than those treated with surgical intervention? The FAME Trial: Health Technology Assessment; NIHR127273.
11. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, 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):1727-36.
12. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware Jr JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. Journal of clinical epidemiology. 2014;67(5):516-26.
13. Resnick B, Jenkins LS. Testing the reliability and validity of the self-efficacy for exercise scale. Nursing research. 2000;49(3):154-9.
14. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American statistical Association. 1996;91(434):444-55.
15. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annual review of public health. 2000;21(1):121-45.
16. Dunn G, Maracy M, Dowrick C, Ayuso-Mateos JL, Dalgard OS, Page H, et al. Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. The British Journal of Psychiatry. 2003;183(4):323-31.
17. Hernández-Alava M, Pudney S. eq5dmap: a command for mapping between EQ-5D-3L and EQ-5D-5L. The Stata Journal. 2018;18(2):395-415.
18. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj. 2009;338.
19. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. BMC Medical Research Methodology. 2013;13:58 -

20. Flight L, Allison A, Dimairo M, Lee E, Mandefield L, Walters SJ. Recommendations for the analysis of individually randomised controlled trials with clustering in one arm–a case of continuous outcomes. BMC Medical Research Methodology. 2016;16:1-13.