OsteoPorosis Tailored Exercise Adherence INtervention trial (OPTIN)

Physiotherapy exercise rehabilitation with tailored exercise adherence support for people with osteoporosis and vertebral fractures:
A randomised controlled trial.

Statistical Analysis Plan

Version 1.2 03 November 2023

Based on protocol version 4.0, 4th March 2022

Registration: ISRCTN:14465704

Role	Name	Title	Signature	Date
Statistical analysis, Trial Manager	Meredith Newman	**Author/Analysis**		24/11/23
Sub study statistical analysis	Erin Hannink	**Author/ Analysis**		24/11/23
Trial Statistician	Ruth Knight	**Author/ Reviewer** **Analysis Oversight**		30/11/23
Chief Investigator	Karen Barker	**Reviewer/Approver**		27.11.23

Contents

1. Introduction

This document details the proposed presentation and analysis for the main paper(s) reporting results from the Chartered Society of Physiotherapy (CSP) Charitable Trust funded OPTIN trial; a prospective, multicentre, individually randomised, controlled, superiority trial with blinded outcome assessment. This Statistical Analysis Plan (SAP) will not detail the analysis approach to two of the sub-studies of the OPTIN trial, the qualitative and health economic analyses. The approach to the analysis of the minimal clinically important difference (MCID) of two key outcomes is included Appendix A.

The results reported in the main trial papers should follow the strategy below. Subsequent analyses of a more exploratory nature will not be bound by this strategy but 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 the published guidelines regarding the content of statistical analysis plans for clinical trials [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 source of the suggestion will be acknowledged.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis will be carried out by an identified, experienced lead member of the research team (MN) and supervised by an appropriately qualified and experienced statistician (RK), who will ensure the integrity of the data during their processing. The MCID analysis will be carried out by EH. Examples of such procedures include quality control and evaluation procedures.

1.1 Key Personnel

Chief Investigator: Professor Karen Barker

Consultant Physiotherapist and Head of Physiotherapy
Nuffield Orthopaedic Centre
Oxford University Hospitals NHS Trust
Email: karen.barker@ouh.nhs.uk
Telephone: 01865 738080
Professor of Physiotherapy, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences NDORMS – University of Oxford

Research Physiotherapist: Erin Hannink

Research Physiotherapist: sub-study analysis
Physiotherapy Research Unit
Oxford University Hospitals NHS FT
Nuffield Orthopaedic Centre, Windmill Rd, Oxford, OX3 7LD

Trial Statistician: Ruth Knight

Statistical Lead
Liverpool Clinical Trials Centre
University of Liverpool
Email: ruth.knight@liverpool.ac.uk

Trial manager: Meredith Newman

Research Physiotherapist, Trial Manager and trial statistical analysis
Physiotherapy Research Unit
Oxford University Hospitals NHS FT
Nuffield Orthopaedic Centre, Windmill Rd, Oxford, OX3 7LD.
meredith.newman@ouh.nhs.uk

1.2 Abbreviations

AE	Adverse event
CI	Confidence Interval
CRF	Case Report Form
EARS-B	Exercise Adherence Rating Scale - part B
EQ-5D-5L	EuroQoL-5 dimensions- 5 level questionnaire
FES-I	Falls Efficacy Scale International
FR	Functional Reach Test
GRC	Global Rating of Change Scale
IQR	Interquartile range
ITT	Intention to treat
MCID	Minimal Clinically Important Difference
NHS	National Health Service
NDORMS	Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
NPRS	Numeric Pain Rating Scale
OPTIN	OsteoPorosis Tailored exercise adherence INtervention
PEQ	Personalised Exercise Questionnaire
QUALEFFO-41	Quality of life for Osteoporosis Questionnaire-41
QoL	Quality of life
RCT	Randomised controlled trial
R&D	NHS Trust Research and Development Department
SAE	Serious Adverse Event

SEE	Self-Efficacy for Exercise scale
6MW	Six-minute walk test
SD	Standard Deviation
SAP	Statistical Analysis Plan
TUG	Timed Up and Go test
TLS	Timed Loaded Standing test
TRC	Trial Research Clinician
VFF	Vertebral Fragility Fracture
wFCI	Weighted Functional Comorbidity Index

2. Background and Objectives

2.1 Background

Vertebral fragility fractures (VFFs) affect at least 20% of older adults in the UK and present a significant health and socioeconomic burden.[2] They are associated with osteoporosis, a progressive systemic skeletal disorder characterised by low bone mass and deterioration in bone microarchitecture that increases bone fragility and fracture risk. VFFs can be difficult to detect, some are initially asymptomatic, but all are associated with an increased risk of mortality in the year post fracture and significantly increase the risk of further fractures.[2, 3] Symptoms become more likely as the number of fractures increase and include back pain, low mood, fatigue, functional disability and decreased quality of life (QoL) which can persist long term.[2-4] Vertebral fractures cause spinal deformity e.g., height loss and excess spinal curvature or thoracic hyperkyphosis. This can restrict pulmonary function, cause abdominal problems, and increase the risk of sustaining further vertebral fractures and of balance deficits, falls and non-vertebral fractures.[2-4] Without treatment, progression and functional decline are expected.[2]

Conservative treatment for osteoporosis includes bone protective medications and lifestyle adaptations.[3] Guidelines recommend people with osteoporosis keep active and exercise to slow the rate of bone loss, to maintain muscle strength and physical function and to prevent falls and further fractures.[2-5] Multiple randomized controlled trials (RCTs) have investigated the efficacy of exercise for people with VFFs and reported short–term, post-treatment benefits to pain, strength, spinal posture, balance, fear-of-falling, mobility and QoL. [4-6] A Cochrane systematic review in 2019 concluded there is moderate-quality evidence that exercise improves physical function.[4] However, treatment effects are often small and difficult to sustain. [4, 6]

Adherence, or the extent to which patients engage in treatment, has been identified as an important issue by many studies, with adherence to clinic-based exercise protocols often around 50% and lower for unsupervised home exercise. [4-7] Partial or non-adherence is associated with worse outcomes and conversely higher adherence with better outcomes. [4, 6, 7] Adherence is a critical consideration because it affects exercise dose. For example, in the PROVE trial benefits associated with exercise at 4 months were not sustained at 12 months, whilst those who attended more sessions experienced better outcomes.[6] Higher adherence has been associated with greater treatment effects but studies monitoring adherence show exercise compliance declines after supervised interventions cease. [7, 8] When assessing intervention efficacy, it is also important to know the minimal clinically important difference (MCID) of any outcome measure in the population concerned, which is defined as the smallest difference in score which a patient perceives as beneficial.[9] A recent Cochrane review identified the Quality-of-Life Questionnaire (QUALEFFO-41) and the Timed Up and Go (TUG) test as important outcome measures targeted by exercise interventions in people with VFFs and also recognised that their MCIDs were unknown.[4]

2.2 Hypothesis and Study Objectives

Hypothesis: Interventions that support patients in adhering to exercise during treatment and for longer self-management on completion of their episode of physiotherapy exercise rehabilitation will increase the efficacy of exercise rehabilitation and benefit older adults with VFF and back pain.

Table 1: Primary and secondary objectives and endpoints

	Objectives	Outcome Measures and Endpoints
Primary	To compare physiotherapy exercise rehabilitation with adherence support with physiotherapy exercise rehabilitation alone in terms of its effects on lower limb strength, walking and balance.	Primary endpoint, 12 months post randomisation. Primary outcome: • Timed Up and Go (TUG) test of lower limb strength, walking and balance (seconds).[10]
Secondary	To compare physiotherapy exercise rehabilitation with adherence support with physiotherapy exercise rehabilitation alone in terms of its effects on: Physical function (balance, strength, pain, spinal curvature, mobility) Quality of life	Secondary endpoints 4, 8 months post-randomisation Secondary outcomes: • QUALEFFO-41: disease specific, self-report QoL questionnaire, 5 sub-scales and total score, each normalised (0-100). Higher better QoL [11] • Falls Efficacy Scale International (FES-I): self-report falls concern, 16 items, each 4 points, total (16-64), higher score more concern.[12] • Numeric pain rating scale (NPRS): self-report back pain intensity (0–10) higher greater pain.

	Fear of falling Exercise self-efficacy Exercise adherence	• Timed Loaded Standing (TLS) test of back muscle endurance, timed (seconds) [13] • Functional Reach (FR): dynamic standing balance test, maximum distance reached (cm).[14] • 6-minute walk (6MW); functional walking capacity distance walked in 6 minutes (metres).[15] • Grip strength (kg); measured with hand-held dynamometer, mean 3 trials each hand.[16] • Thoracic kyphosis; measured with a flexicurve, a non-radiological measure of spinal curve (degrees).[17] • Falls number/ nature/ outcome: prospective log. • Exercise self-efficacy scale (SEE) – 9 items, each 10 points, scores summed (0-90). Higher scores = greater self-efficacy. [18] • Exercise adherence: Exercise Adherence Rating Scale (EARS), 6 items, each 5 points, higher more adherence. Session attendance. [19] • Adverse events: number/ nature prospective log. Healthcare Diary: prospective log of health professional and hospital visits
Exploratory	To determine the minimally important clinical difference (MCID) of the TUG test and the QUALEFFO-41 after physiotherapy in patients with VFFs and back pain.	At 4 months post randomisation • TUG • QUALEFFO-41 • TLS • EQ-5D-5L, generic measure of health related QoL, 5 domains, 5-point. [20] • Global rating of change (GRC) scale: patient perspective of change rated on 7-point ordinal scale, higher scores more positive change. [21]

3. Study Methods

3.1 Study Design

OPTIN is a two-arm, parallel group, superiority randomised controlled trial (RCT) with blinded assessments. It includes a nested evaluation of the MCID of key outcome measures. All patients receive treatment, either a standardised package of physiotherapy exercise rehabilitation or the same exercise rehabilitation integrated with an exercise behaviour change intervention (OPTIN intervention). The primary outcome is the TUG measured at 12 months post-randomisation. All clinical outcomes are collected at baseline (0) and 4-, 8- and 12-months following randomisation.

The trial will take place in at least 6 NHS hospitals across the UK and their related physiotherapy services and via video-call in participant homes. Each site will recruit participants, undertake measures and provide the intervention and comparator. A trial research clinician (TRC) will collect

data at research assessments, following a standardised format and using outcome measures that are valid and reliable with this patient population. Measures are a mix of self-report questionnaires and measures of physical function, including monitoring adverse events and intervention adherence.

3.2 Eligibility

Participants will be adults aged 55 years or over who have a diagnosis of at least one vertebral fragility fracture due to osteoporosis and back pain who are willing and able to give informed consent. Specifically:

Inclusion Criteria

- Men and women ≥ 55 years: all women at least 1 year post-menopausal
- One or more VFFs confirmed by radiography, X-Ray, MRI or DEXA scan, people with VFF of any severity and at any time-point post-fracture are eligible
- Back pain in the previous 12 months
- Able to walk at least 10 metres independently with or without a walking aid

Exclusion Criteria

- Conditions that would make participating in physiotherapy or exercise unsafe or confound results e.g., significant neurological and psychiatric conditions, severe unstable cardiovascular or pulmonary disease
- Bone loss secondary to other metabolic disorders, diseases or medications e.g., rheumatoid arthritis, anorexia
- Primary problem is back pain that involves pain radiating into the lower limbs

3.3 Sample size

The primary outcome is the TUG test, a test of balance, lower limb strength and walking ability with established reliability and validity. It records the time in seconds (s) a person takes to stand up from a chair, walk 3 metres at a self-selected speed, turn and walk back and sit down, faster performance indicates better function. [10, 15] The MCID for the TUG has not been established in people with VFFs, but a MCID of 1.4 s is reported for similar older populations with chronic musculoskeletal disorders.[22] The study requires 104 participants (52 per arm) to be 80% powered to detect a 1.4s difference in TUG score between groups at a 5% significance level (two sided) assuming that the SD is 2.5s. Similar trials have had loss to follow-up rates of 10% at 12 months. To account for this the sample size is 116 participants (58 per arm).

Sample size calculations were performed in Stata v15 (StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC).

3.4 Randomisation and blinding

Following screening and confirmation of eligibility, baseline assessment and registration participants will be individually randomised (1:1) between either: OPTIN intervention or comparator

(physiotherapy exercise rehabilitation) trial arms. The computer-generated randomisation schedule will be prepared by the trial statistician (RK). Individual randomisation will be stratified by recruitment centre and permuted blocks of varying undisclosed sizes will be used. Full details of randomisation, concealment and allocation procedures are available in the OPTIN protocol version 4.0, 4th March 2022, stored in the trial master file (TMF.)

The TRC and study personnel conducting measures and entering data will be blinded to allocation group. Due to its nature as an exercise and behaviour change intervention, participants and those physiotherapists delivering the rehabilitation will be aware of treatment allocation and hence a code-breaking facility is not required. To reduce contamination, treating physiotherapists will be allocated to one arm of the trial, provided with the training and resources for this arm alone, and only deliver treatments to participants allocated to this arm.

3.5 Treatment Interventions per Allocated Arm

Table 2: Summary of Intervention and Comparator Treatments

OPTIN Intervention	Standardised physiotherapy exercise rehabilitation: • Physiotherapy musculoskeletal assessment (60 mins) • Six follow-up sessions (30 mins each) over 16 weeks Extended assessment (additional 30 mins) comprising Personalised Exercise Questionnaire (PEQ); self-report questionnaire about exercise behaviours and extended interview to assess capability (C), opportunity (O) and motivation (M) regarding exercise behaviours (B), (COM-B). (Integrated with 60 min assessment above) Individualised prescription of techniques to support exercise behaviour. At least 3 techniques, selected by physiotherapist from a toolkit of 9 recognised techniques (60 mins total, spread over 16 weeks).
Comparator	Standardised physiotherapy exercise rehabilitation: • Physiotherapy musculoskeletal assessment (60 mins) • Six follow-up sessions (30 mins each) over 16 weeks

3.6 Outcome Assessment Schedule

At the baseline assessment information is collected about relevant demographic characteristics and past medical history including fragility fractures, falls in previous year, bone mineral density (BMD) and co-morbid conditions using the weighted Functional Co-morbidities Index. [23] The participant is then asked to complete the outcome measure package, this is then completed again at 4, 8 and 12 months plus the adherence (EARS-B: 4, 8, 12 months) and global rating of change (GRC: 4 months) scales which can only be completed post-treatment. Session attendance is logged by physiotherapists across the intervention period.

Table 3: Outcome Measure Time-points

MEASURES	Month-0	Month-4	Month-8	Month-12

Falls in past year	✓			
Falls (prospective log)		✓	✓	✓
NPRS (back pain last 2 weeks)	✓	✓	✓	✓
Thoracic kyphosis	✓	✓	✓	✓
TUG	✓	✓	✓	✓
FR	✓	✓	✓	✓
TLS	✓	✓	✓	✓
QUALEFFO-41	✓	✓	✓	✓
SEE	✓	✓	✓	✓
FES-I	✓	✓	✓	✓
EQ-5D-5L	✓	✓	✓	✓
EARS-B		✓	✓	✓
6 MW	✓	✓	✓	✓
Grip strength	✓	✓	✓	✓
Healthcare use (prospective log)		✓	✓	✓
Adverse Events		✓	✓	✓
GRC		✓		

3.7 Quality Control and Data Validation

Data will be collected from participants and the research team onto paper questionnaires/CRFs at research assessment visits. The originals will be scanned, and then sent by a member of the local research team to the study coordinating office in Oxford by post, using a Freepost account, keeping a copy at site and sending a copy electronically via secure NHS networks. Upon receipt of questionnaires/CRFs, appropriate data quality and validation checks will be carried out by the study research administrator and the data entered into a study-dedicated databases.

3.8 Statistical Analysis Outline

It is anticipated that all statistical analysis will be undertaken using Stata (StataCorp LP, www.stata.com) or other well-validated statistical packages such as SPSS. A blinded analysis of the data (not separated by treatment group) will be undertaken prior to the final data lock to look into the distribution of variables, missing data distributions, and to finalise the per protocol population.

Standard descriptive statistics will be used. Means and standard deviations (SDs) or medians (with lower and upper quartiles) as appropriate will be used for continuous variables, and numbers and percentages will be used for binary and categorical variables. Summary statistics will be presented for all comparative outcomes, and effect estimates will be reported together with 95% confidence intervals with all tests carried out at a 5% two-sided significance level.

The final analysis of all primary and secondary endpoints will be conducted together when all recruited patients have completed all follow-up. The study sample will be described, and then principal comparisons will be performed on an intention-to-treat basis, with a per-protocol analysis repeated at 12 months for the primary outcome, the Timed up and Go (TUG) test.

At 12 months post-randomisation the two treatment groups will be compared on the TUG measure using a multivariate linear regression model adjusting for recruiting centre (stratification factor), age and baseline TUG score. An unadjusted t-test will also be undertaken. An additional analysis utilising all time points will be undertaken. The assumption of approximate normality will be assessed by examining the residuals. If this assumption is not met the first approach will be to consider a transformation to achieve normality. If this is not possible, the two groups will be compared using non-parametric methods (e.g., Mann-Whitney U-test).

Similar analyses will be performed for secondary outcomes which can be considered approximately continuous (QUALEFFO-41, FES-1, NPRS, TLS, Grip strength, 6MW, FR, Thoracic kyphosis, SEE and EARS) at 4-, 8- and 12-months post-randomisation. The appropriateness of the assumption of approximate normality will also be considered and transformation to normality or non-parametric methods used as appropriate. It is not anticipated that the data about healthcare visits or falls will be approximately normally distributed, therefore, this data will be summarised by treatment group using medians and IQRs and compared using non-parametric methods. Data about falls will also be included in safety data reporting e.g., fall nature and severity (see section 7 below).

4. Statistical Principles

4.1 Statistical Significance and Multiple Testing

A two-sided 5% significance level will be used throughout with 95% CIs presented. Since there is a single primary outcome in this study there is no concern regarding multiple testing, and all secondary outcomes will be considered as supporting the primary result.

4.2 Definition of Populations for Analysis

All analyses will be performed for the intention to treat (ITT) population. This will include all randomised participants with available data who will be analysed according to their allocated intervention regardless of the treatment they received.

In addition, analysis of the primary outcome (TUG at 12 months) will be repeated for the per protocol (PP) population which will include only those participants who received their allocated treatment, that is the participants in the intervention arm that receive the OPTIN assessment interview and at least 3 adherence techniques and the participants in the comparator arm that do not. Participants with other major protocol deviations (e.g., recruited and later found to be ineligible) will also be excluded from this population.

4.3 Procedure for Accounting for Missing, Unused and Spurious Data

Missing data, e.g., due to withdrawal, protocol deviation or patient loss to follow-up will be summarised and patterns analysed. The proportion of missing values per variable will be presented. Where appropriate, differentiation will be made between partially completed and fully missing outcome data. For specific questionnaires their established, validated methods for managing missing item-level data will be adopted e.g., QUALEFFO-41. The main analyses will be performed using available cases only. If needed, a sensitivity analysis of the primary outcome (TUG at 12 months post-randomisation) will explore the effect of missing data on the primary outcome results using multiple imputation techniques. These will explore the possibility of data being missing at random as well as potential departures from this assumption if appropriate.

Table 4. Summary of missing data at each point

	Intervention			**Control**			**Total**		
	Expected n	Received n %	Missing from Expected n %	Expected n	Received n	Missing from Expected n %	Expected n	Received n	Missing from Expected n %
Baseline demographic									
TUG Baseline 4-months 8-months 12-months									
Back pain									

Baseline 4-months 8-months 12-months									
6MW Baseline 4-months 8-months 12-months									
FR Baseline 4-months 8-months 12-months									
TLS Baseline 4-months 8-months 12-months									
QUALEFFO-41 Baseline 4-months 8-months 12-months									
Grip *(dominant hand)* Baseline 4-months 8-months 12-months									
FES-I Baseline 4-months 8-months 12-months									
SEE Baseline 4-months 8-months 12-months									
Kyphosis									

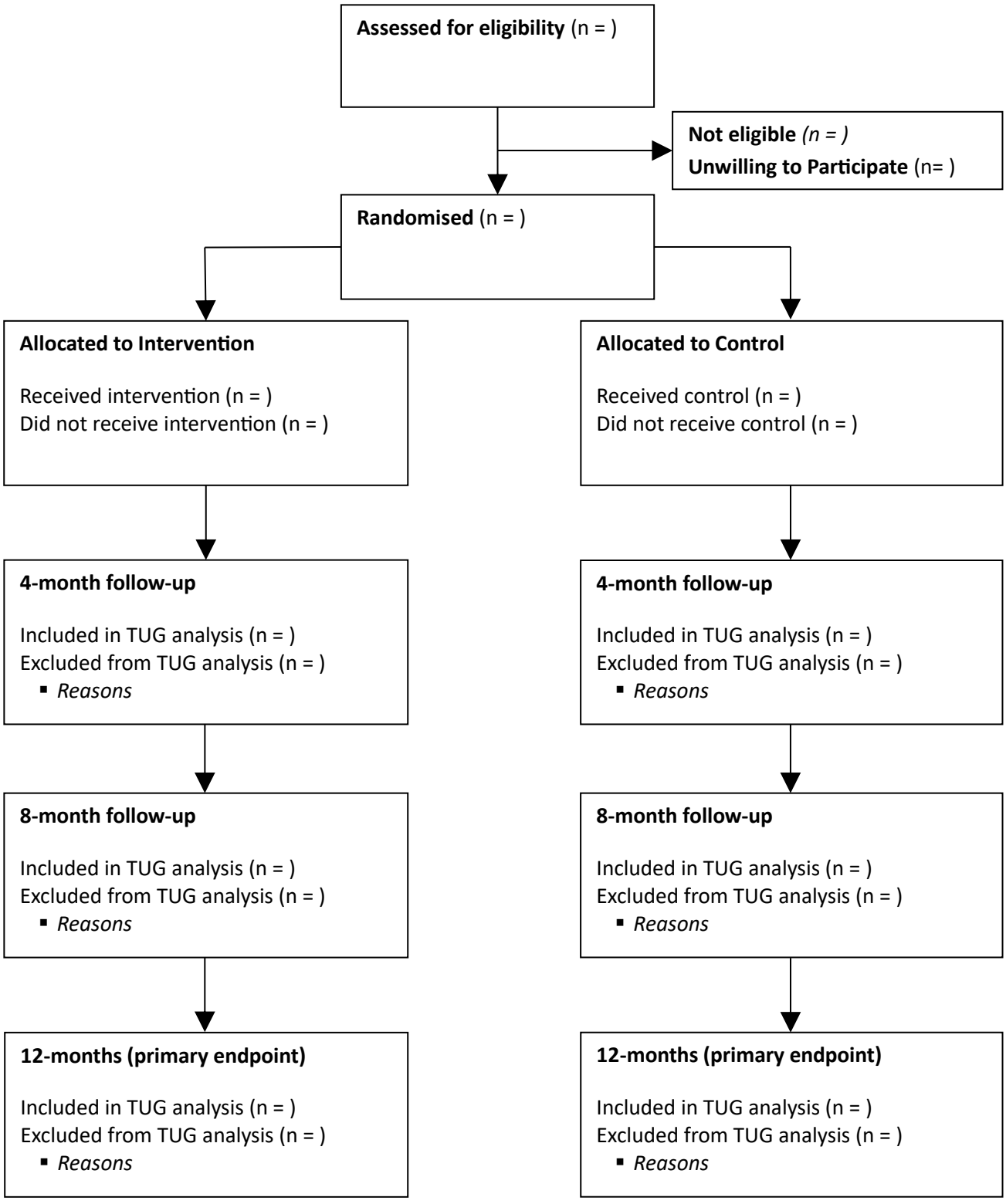
Baseline 4-months 8-months 12-months									
EARS 4-months 8-months 12-months									
GRC 4-months									
EQ-5D-5L Baseline 4-months 8-months 12-months									
Falls log 4-months 8-months 12-months									
Healthcare diary 4-months 8-months 12-months									

5. Trial Population and Descriptive Analyses

5.1 Participant flow and representativeness of study sample

The flow of participants through the trial from screening (number screened, meeting eligibility, willing to participate), randomisation, receiving treatment, to follow-up and inclusion in primary analyses will be summarised using a CONSORT flowchart (Figure 1). Reasons [24] Numbers of participants recruited and included and excluded in final analyses for the primary outcome will be displayed.

Figure 1: CONSORT flow diagram



5.2 Withdrawal from treatment and loss to follow-up

The number and proportion of participants who withdraw along with timings of withdrawal from the trial, as well as the number of lost to follow up or excluded from the primary analysis will be summarised along with reasons in Table 5. Then used to inform Figure 1. If differential losses are identified, the reasons for these will be explored further.

Table 5. Loss to follow-up and withdrawal

Time	Intervention	Control	Total, n (%)
Prior to treatment Withdrawn, n Loss to follow up, n Total, n (%)			
Before follow-Up 1 (4-month) Withdrawn, n Loss to follow up, n Total, n (%)			
Between follow-Up 1- 2 (8-month) Withdrawn, n Loss to follow up, n Total, n (%)			
Between follow-Up 2-3 (12-month) Withdrawn, n Loss to follow up, n Total, n (%)			
Total, n (%)			

Table 6. Reasons withdrawn or loss to follow up

Reason	Intervention	Control	Total, n (%)
Consent withdrawn			
SAE/AE/Medical conditions[a]			
Uncontactable			
…			

Total			

a. See Table 16: Adverse events and Serious Adverse events

5.3 Description of Compliance with Intervention

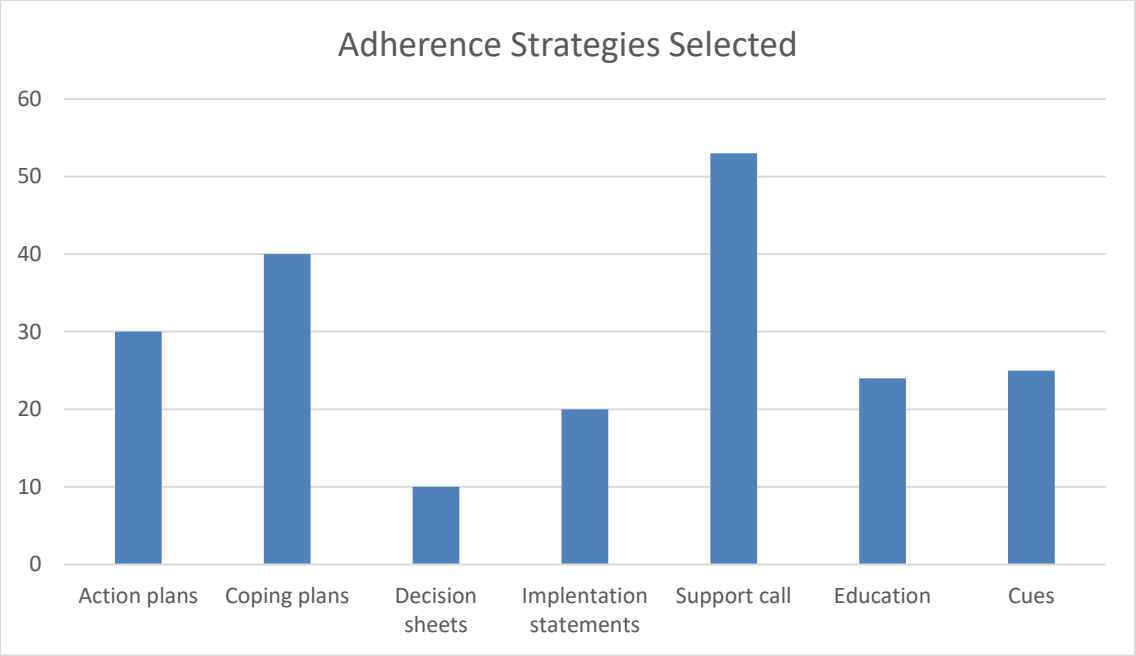
Treatment fidelity will be reported. The total number of trial physiotherapy sessions attended, and the number of these that were delivered virtually will be summarised for each group. Compliance with the intervention is defined as participating in the assessment and prescription of at least 3 adherence support techniques as recorded on treatment logs. Treatment session numbers and as well as the adherence strategies prescribed for the intervention group. We will display graphically the types and frequency with which adherence strategies were selected (Figure 2). We will summarise participant visits to healthcare professionals outside of the OPTIN trial, and those related to back pain to monitor total dose of physiotherapy (private and NHS) and back-related intervention, see Table 8.

Table 7: OPTIN sessions attended and compliance.

	Intervention			**Control**	
	Physiotherapy Session n	**Virtual session** n	**Behavioural Techniques** n	**Physiotherapy Session** n	**Virtual session** n
Assessment*					
Follow-up 1					
Follow-up 2					
Follow-up 3					
Follow-up 4					
Follow-up 5					
Follow-up 6					
Total (median IQR)					
Compliant n (%) **Assessment and ≥ 3 techniques**					

*Extended 90 min for OPTIN arm. Virtual session = physiotherapy via telephone or video-call.

Figure 2. Adherence strategies selected (example)



Information about health care received outside of the trial i.e., number of health care visits, reason for visit (back related/ un-related), health professional seen (GP/Nurse/Physiotherapist/OT/Other), hospital stays is not expected to be normally distributed. Total visit numbers, number of physiotherapists visits, (private and NHS) reason for visit (back related/ unrelated) will be summarised for each group.

Table 8: Healthcare and non-trial physiotherapy appointments

	Total n (%) median (IQR)	Intervention n (%) median (IQR)	Control n (%) median (IQR)
Number of health professional appointments			

Number of back-related health professional appointments			
Number of physiotherapist appointments			
Private physiotherapist appointments			
NHS physiotherapist appointments			

5.4 Unblinding

This is an assessor only blinded trial. If trial assessors become unblinded to treatment group, they will treat this as a protocol deviation. Data about protocol deviations will be summarised and reported as part of safety reporting: see section 7.

5.5 Baseline Comparability of Randomised Groups

Standard descriptive statistics will be used to describe the characteristics of the two groups at baseline that is numbers with proportions for binary or categorical variables and either means (with SDs) or medians (with IQRs) for continuous variables as appropriate. Baseline comparability of the groups will be summarised including demographic characteristics and performance on primary and secondary outcomes as well as on other important prognostic factors e.g., co-morbidities, number of fragility fractures. In addition, thoracic kyphosis at baseline will be plotted for each group to better understand the distribution of spinal curvatures in the study population i.e., hypokyphotic (lower curvature), normal (mid) ranges of thoracic kyphosis, to moderate or severe hyperkyphosis (higher curvature). There will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any baseline variable.

Table 9: Demographic and baseline characteristics of participants

	Intervention (n = XX)	**Control** (n = XX)	**All participants** (n = XX)
Sex [a]			
Male			
Female			

Age [b] (years)			
Height [b] (cm)			
Weight [b] (kg)			
BMI [b] (kg/m^2)			
Comorbidities score (wFCI) [b]			
BMD (g/cm^3) [b]			
T-score lumbar spine [b]			
Time since osteoporosis diagnosis [b]			
Number of vertebral fractures [b]			
Location of vertebral fractures [a] Upper thoracic (T1-T5) Lower thoracic (T6-T12) Upper Lumbar (L1, L2) Lower Lumbar (L3-L5)			
Last confirmed VFF [b] **(months)**			
Number of non-vertebral fractures [b]			
Type of non-vertebral fractures [a] Shoulder Wrist Hip Other Fragility Fractures			
Self-report mobility (walking) [a] 1. Unlimited 2. 500m-1km 3. 100-500m 4. <100m 5. Housebound 6. Unable			

Self-report mobility (stairs) [a] 1. Normal 2. One step at a time 3. Down with rail 4. Up and down with rail 5. Unable down 6. Unable			
Self-report mobility (support) [a] 1. None 2. Stick outdoors 3. Stick always 4. 2 sticks 5. 2 crutches 6. Walking frame			

a: binary and categorical data: n (%), b: continuous data: mean (SD) or median (25-75% quartiles), in either case the number of participants providing data will also be reported

Table 10: Baseline primary and secondary outcome measures

	Intervention (n = XX)	**Control** (n = XX)	**All participants** (n = XX)
Presence of back pain [a] In past 2 weeks Today			
Back pain intensity (NPRS 0-10) [b] In past 2 weeks Today			
Fall in past year [b] Needed medical attention			
Participants falling [a] 1.Frequently (≥ 1/ week) 2.Occasionally (≤ 1/ month) 3. Rarely (≤ 1/ year)			
TUG (s) [b]			
QUALEFFO-41 Total (points) [b]			

QUALEFFO-41 subscales (points) [b] Pain Physical Function Social Function General Health Mental Health			
Thoracic kyphosis (degrees) [b]			
Grip dominant hand (kg) [b]			
6MW (m) [b]			
6MW exertion CR10-RPE (0-10) [b] Legs Breathing			
FR (cm) [b]			
TLS (s) [b]			
FES-I [b]			
SEE [b]			

a: binary and categorical data: n (%), b: continuous data: mean (SD), range or median (25-75% quartiles), range in either case the number of participants providing data will also be reported

5.6 Reliability

To ensure consistency and reliability of data processing, validation checks will be conducted. This will include checking for duplicate records, checking for missing data, checking the validated values for implausible values and validating potential outliers, and checking that data has been imported into statistical packages correctly. This will involve records from each site. Calculations and processes performed by computer e.g., QUALEFFO-41 calculations will be checked by hand calculations. These checks will be performed for 25 participants randomly sampled from the dataset. Clarification will be sought by comparing electronic datasets with case report forms.

6. Analysis

6.1 Methods used for analysis of primary outcome.

We will plot and display TUG data at all time points: 0, 4, 8 and 12-months to consider patterns of change over time and any outliers. At 12-months post-randomisation the two treatment groups will be compared on the primary outcome of the TUG measure using a multivariate linear regression model adjusting for recruiting centre (stratification factor), age at baseline and baseline TUG score. An unadjusted t-test will also be undertaken. As the TUG is recorded at 4 months and 8 months after randomisation as well, an additional analysis utilising all time points using multi-level modelling and including a treatment by time interaction, if appropriate, will be undertaken. For each of these models, the assumption of approximate normality will be assessed by examining the residuals. If this assumption is not met the first approach will be to consider a transformation to achieve normality. If this is not possible, the two groups will be compared using non-parametric methods (e.g., Mann-Whitney U-test). This analysis will be unadjusted and will consider each time point separately.

Table 11: Analysis of primary endpoint, TUG test change at 12-months

TUG (seconds)*	**Intervention** Mean (SD)	**Control** Mean (SD)	**Mean difference** (95% CI)	p-value
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting analyses				
Repeated measures				
4 months				
8 months				
12 months				
12 months (per protocol)				

* A negative change corresponds to a better result at 12-months

6.2 Methods used for analysis of secondary outcomes.

Secondary outcomes which can be considered approximately continuous (QUALEFFO-41, FES-1, NPRS, TLS, Grip strength, 6MW, FRT, thoracic kyphosis, SEE and EARS) will be plotted over time and compared at 12 months using a multivariate linear regression model adjusting for recruiting centre (stratification factor), age at baseline and baseline score of the relevant variable as appropriate e.g. for the QUALEFFO-41 this would be baseline QUALEFFO-41. An unadjusted t-test will also be completed. An additional analysis using data at 4-, 8- and 12-months post-randomisation will also be undertaken using a repeated measures linear regression model. The appropriateness of the assumption of approximate normality will also be considered and transformation to normality or non-parametric methods used as appropriate. See Table 12.

Thoracic kyphosis is unlikely to be normally distributed, as in populations with vertebral fracture more people are likely to have moderate and severe kyphosis. [2, 25]. In addition, previous work has suggested the potential for change in thoracic kyphosis at follow-up is greater for those with the highest baseline values of kyphosis. [25] Therefore we will plot and display thoracic kyphosis values and consider a transformation to normality before comparison using parametric tests. If normality is not possible, non-parametric methods will be used.

Table 12: Comparison between groups secondary outcomes, self-report questionnaires

	Intervention Mean (SD) n	**Control** Mean (SD) n	**Mean Difference** (95% CI)	p-value
QUALEFFO-41 (total points)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
QUALEFFO-41 (pain)				
Unadjusted 12-months				

Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
QUALEFFO-41 (physical)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
QUALEFFO-41 (social)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
QUALEFFO-41 (general)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				

4 months				
8 months				
12 months				
QUALEFFO-41 (mental)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
FES-I				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
NPRS today (0-10)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
NPRS last 2 weeks (0-10)				

Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
SEE				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
EARS-B				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				

Table 13: Comparison between secondary outcomes, physical measures

	Intervention Mean (SD) n	**Control** Mean (SD) n	**Mean Difference** (95% CI)	p-value
TLS (s)				

Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
Grip strength dominant hand (kg)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
6MW (m)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				
Functional Reach Test (cm)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				

Supporting Analyses				
4 months				
8 months				
12 months				
Thoracic Kyphosis (Degrees)				
Unadjusted 12-months				
Adjusted for baseline, centre and age				
Supporting Analyses				
4 months				
8 months				
12 months				

Falls are both a secondary outcome and important safety data. The nature, severity and outcome of falls will be reported as part of trial safety adverse event monitoring: see section 7 below. In addition, numbers and additional data about falls collected at baseline and from fall diaries will be reported and analysed. The number and proportion of participants who have a fall during follow-up will be summarised by treatment group and compared between the groups using a logistic regression model. This model will be adjusted for recruiting centre, age at baseline and whether or not the participant reported a fall in the previous year at baseline. Odds ratios will be reported along with 95% CIs and associated p-values. Amongst those participants who report at least one fall, the number of falls reported during follow-up will be summarised by treatment group using medians and IQRs and compared using a Mann-Whitney U-test. This will be repeated for falls that require medical attention. These analyses will be presented as outlined in Table 13. If a sufficient number of falls are reported this analysis may be reported considering each portion of follow-up (i.e., up to 4 months, from 4 to 8 months and from 8 to 12 months) separately.

Table 14: Comparison between groups falls data.

	Intervention	Control	OR (95% CI)	p-value
Number of participants reporting a fall [a]				
Number of falls reported [b]			NA	

Number of falls needing medical attention [b]			NA	
Number of falls in participants that previous reported falling in previous year at baseline [b]				

[a] Summaries will be n (%); [b] Summaries will be median (IQR)

7.0 Safety Reporting

Trial protocol deviations which have been recorded for each centre, with reasons and participant trial status outcomes, will be tabulated: see Table 16. Then the number, impact of protocol deviation and proportion of participants experiencing a protocol deviation will be summarised for the trial as a whole and per treatment group: see Table 17.

Table 16: Protocol Deviation.

Number	Nature of Deviation	Site	Trial Arm	Deviation Impact	Trial Status
1					
2					
…					
Total					

Trial Arm: Intervention or Control, Deviation Impact: 1=Completeness of trial data, 2= Reliability of trial data, 3=Primary outcome, 4= Accuracy of trial data, 5= participant rights, safety, 6=no impact

Table 17: Summary Table of Protocol Deviations

	Total n	Intervention n (%)	Control n (%)

Number of Protocol Deviations			
Number of participants with a protocol Deviation			
Deviation of Impact 1.Completeness 2.Reliability 3.Primary outcome 4.Accuracy 5.Participant 6.No impact			
Total	NA		

All AEs and SAEs will be described, including their nature, severity, relatedness to the intervention and outcome on participant trial status; See Table 18 and 19. This includes details such as any unplanned hospital admission connected to adverse events. The number and proportion of participants experiencing an AE or SAE during the trial will be summarised by treatment group. Depending on the amount of data, a logistic regression model adjusted for recruiting centre will be used to compare the rates in the two groups. Deaths are not anticipated in this study, but details of any that do occur will be reported.

Table 18: Adverse Events.

Number	Nature of Event	Severity	Status of Event	Causality (Related or Unrelated)	Trial Status
1					

2					
…					
Total					

Severity 1=mild, 2=moderate, 3=severe, 4= not known. Status: 1=resolved, 2=recovered with sequalae, 3=ongoing, 4=not known. Causality: Related=1, unrelated=0.

Table 19: Serious Adverse events.

Number	Nature of Event	Status Of Event	Causality (Related or Unrelated)	Hospital stay * (days)	Trial Status
1					
2					
…					
Total					

Status: 1= resolved, 2=recovered with sequalae, 3=ongoing, 4= fatal, 5= Not known. Causality: related=1, unrelated=0. *Hospital admission if applicable/ data available.

Table 20: Comparison of adverse events between groups.

	Total n	**Intervention** n (%)	**Control** n (%)	**OR (95% CI)***	**p-value**
Number of SAEs					
Number of participants with serious adverse events					
Number of AEs					

Number of participants with adverse events					
Number of related AE events					
Number of related SAE events					
Total					

- Comparison dependent on quantity of data

8. References

1. Gamble, C et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA, 2017.318(23): p. 2337-2343.
2. International Osteoporosis Foundation. Epidemiology of Osteoporosis: International Osteoporosis Foundation. URL https://www.osteoporosis.foundation/health-professionals/fragility-fractures/epidemiology (accessed 12 August 2023).
3. NOGG 2021: Clinical guideline for the prevention and treatment of osteoporosis. https://www.nogg.org.uk/full-guideline/section-8-management-symptomatic-osteoporotic-vertebral-fractures (accessed 12 August 2023)
4. Gibbs JC, MacIntyre NJ, Ponzano M, *et al*. Exercise for improving outcomes after osteoporotic vertebral fracture. *Cochrane Database Syst Rev* 2019;7:CD008618.
5. Brooke-Wavell K, Skelton DA, Barker KL. Strong, steady and straight: UK consensus statement on physical activity and exercise for osteoporosis. *Br J Sports Med* 2022;104634:846. doi:10.1007/s00198-019-05133-0
6. Barker K, Newman M, Stallard N. Physiotherapy rehabilitation for osteoporotic vertebral fracture: a randomised controlled trial and economic evaluation (prove trial). *Randomized Controlled Trial* 2019. doi:10.1007/s00198-019-05133-0
7. Falossi F, Azzollini V, Notarstefano C, Raffaetà G. Adherence to a home physical exercise program in patients with osteoporotic vertebral fractures: A retrospective observational study. *Journal of Back and Musculoskeletal Rehabilitation.* 2022; 35 (4): 777-782
8. Room J, Hannink E, Dawes H, *et al*. What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review. *BMJ Open* 2017; 7:e019221.

9. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10 (4):407-415. doi:https://doi.org/10.1016/0197-2456(89)90005-6
10. Podsiadlo, D. and Richardson, S. "The timed "Up & Go": a test of basic functional mobility for frail elderly persons." *J Am Geriatr Soc*. 1991;39 (2):142-148.
11. International Osteoporosis Foundation Quality of life questionnaires QUALEFFO-41. URL: https://www.osteoporosis.foundation/health-professionals/research-tools/quality-life-questionnaires (accessed 12 August 2023).
12. Delbaere K, Close JCT et al. The falls efficacy scale international (FES-I). A comprehensive longitudinal validation study. *Age and Ageing.* 2010; 39: 201-16.
13. Newman M, Newman R, Hughes T, Vadher K, Barker KL. Is the timed loaded standing test a valid measure of back muscle endurance in people with vertebral osteoporosis? Osteoporos Int 2018; 29: 893–905. https://doi.org/10.1007/s00198-017-4358-8
14. Weiner DK, Duncan PW et al. Functional reach: a marker of physical frailty. *J Am Geriatr Soc* 1992; 40(3):203-7.
15. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up and Go Test, and gait speeds. *Phys Ther.* 2002;82:128–37.
16. Abizanda P, Navarro JL et al. Validity and usefulness of hand held- dynamometry for measuring muscle strength in community-dwelling older persons. Arch Gerontol Geriatr. 2012; 54(1): 21-7. DOI: 10.1016/j.archger.2011.02.006.
17. Greendale GA, Nili NS, Huang MH, Seeger L, Karlamangla AS. The reliability and validity of three non-radiological measures of thoracic kyphosis and their relations to the standing radiological Cobb angle. Osteoporos Int 2011;22:1897–905. https://doi.org/10.1007/s00198-010-1422-z
18. Resnick B and Jenkins LS. Testing the reliability and validity of the self-efficacy for exercise scale. Nurs Res 49 (3): 154-159.
19. Newman-Beinart NA et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy 103 (2017) 180-85.
20. Herdman M, Gudex C, Lloyd A et al. Development and preliminary testing of the new five-level version of the EQ-5D-5L (EQ-5D-5L). Qual Life Res 2011; 20 (10): 1727-36.28.
21. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-170. doi:10.1179/jmt.2009.17.3.163
22. Wright A, Cook C, Baxter G, et al. A comparison of 3 methodological approaches to defining major clinically important improvements of 4 performance measures in patients with hip osteoarthritis. *J Orthop Sport Phys*. 2011; 41: 319 -27.
23. Kabboord AD et al The modified functional co-morbidity index performed better than the Charlson index and original functional comorbidity index in predicting functional outcome in geriatric rehabilitation: a prospective observational study. *BMC Geriatrics* 2020 20: 114. https://doi.org/10.1186/s12877-020-1498-z
24. Schulz, K.F., Altman, D.G. and Moher, D., 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*, *11*(1), pp.1-8.
25. Hannink e, Newman M, Barker K. Does thoracic kyphosis severity predict response to physiotherapy rehabilitation in patients with osteoporotic vertebral fracture? A secondary analysis of the PROVE RCT. *Physiotherapy*. 2022; 85-92.

Appendix A.

MCID sub-study statistical analysis plan

The aim of this sub-study is to estimate the MCID of the TUG test and the QUALEFFO-41 questionnaire. In addition to the key outcome measures (TUG and QUALEFFO-41), the following variables will be used as anchors:

TUG test anchors:

1. Global rating change (GRC) scale of walking and balance – self-reported rating at 4 months
2. Timed Loaded Standing (TLS) test – change between scores at 4 months and baseline
3. Physical Function (PF) domain of QUALEFFO-41 – change between scores at 4 months and baseline.

QUALEFFO-41 anchors:

1. GRC scale of quality of life – self-reported rating at 4 months
2. TLS test – change between scores at 4 months and baseline
3. EQ-5D-5L – change between overall health scale at 4 months and baseline

Participants will be included in the analysis if they have complete baseline and 4-month data for target outcomes and their respective anchors variables. For this sub-group we will perform descriptive statistics (mean and standard deviation (SD), median and interquartile range (IQR), or percentage, where appropriate) for sex, age, height, weight, BMI, comorbidities score, BMD, and T-score lumbar spine.

We will then perform multiple Pearson's correlation analyses for the key outcome measures and each of their corresponding anchors. Anchors that do not show at least a moderate correlative relationship ($r>0.4$) with their key outcome measure will not be included in further MCID analyses.

We will calculate the change between 4-month scores and baseline for both key outcome measures and eligible anchor variables, excluding GRC scales as they are self-reported only at 4 months.

From the 4-month GRC scores and the change scores calculated from the other anchor variables, we will dichotomise data into 'responder' or 'non-responder' based on the following:

a) GRC is a 7-point scale: -3 (much worse), -2 (moderately worse), -1 (a little worse), 0 (no change), 1 (a little better), 2 (moderately better), 3 (much better); scores >0 will be labelled as a 'responder' and scores ≤0 labelled as a 'non-responder'.
b) TLS is a strength test where longer duration indicates stronger musculature; calculated change scores >0 seconds will be labelled as a 'responder' and ≤0 seconds labelled as a 'non-responder'.
c) PF domain of the QUALEFFO-41 is a questionnaire where a *lower* score indicates better physical aspects of quality of life; calculated change scores <0 will be labelled as a 'responder' and scores ≥0 labelled as a 'non-responder'.

d) ED5Q overall health scale is reported on a visual analogue scale from 0-100, where 100 indicates best health; calculated change scores >0 will be labelled as a 'responder' and scores ≤0 will be labelled as a 'non-responder'.

Based on these dichotomised data, we will perform anchor-based analyses for the TUG and QUALEFFO-41 using each eligible corresponding anchor variable. Methods of anchor-based analysis will include:

a) Average Change (AC) – the average score change of the responders, according to the anchor.
b) Change Difference (CD) – the average score of change of responders (according to the anchor) minus the average score of change of non-responders.
c) Receiver-operating characteristic (ROC) analysis - a plot of sensitivity against specificity using dichotomised data from each anchor. We will use Youden's J statistic and Euclidean distance to determine the optimal cut-off values, and also calculate the area under the curve (AUC).

In addition to the anchor-based methods, we will estimate the MCID with distribution-based analysis by calculating 0.5SD of the TUG and QUALEFFO-41 change scores (between baseline and 4 months).

Example Table A: MCID values for key outcome measures

Key outcome measure	**Corresponding anchor**	**Method of analysis**			
		AC (s)	CD (s)	ROC (s)	0.5 SD (s)
TUG	GRC				
	PF				
	TLS				
QUALEFFO-41	GRC				
	EQ-5D				
	TLS				

Example Table C: ROC analyses

Key outcome measure	**Corresponding anchor**	**Youden's J statistic**	**Euclidean distance**	**AUC**
TUG	GRC			
	PF			
	TLS			
QUALEFFO-41	GRC			
	EQ-5D			
	TLS			