Data Analysis Plan (DAP)

iFraP: Improving uptake of Fracture Prevention drug Treatments

A person-centred approach to improving uptake of Fracture Prevention drug Treatments: a randomised controlled trial of the iFraP intervention in Fracture Liaison Services

Version 2.0

Date: 31st August 2024

ISRCTN/Clinical trials.gov Number: ISRCTN10606407

This document has been written based on version 1.6. of the study protocol, dated the 13th of December 2023 (https://doi.org/10.1186/ISRCTN10606407). Our study protocol has also been published in NIHR Open Research (https://openresearch.nihr.ac.uk/articles/4-14et al. 2021)

Data Analysis Plan (DAP) revision history

Protocol version	Updated DAP version	Section number changed	Description of and reason for the change	Date changed

Roles and responsibilities

The undersigned have written the data analysis plan for the iFraP trial and agree its content:

Name	Role	Signature	Date
Dr Elaine Nicholls	Statistician	[redacted]	31/08/2024
Dr Laurna Bullock	Associate Investigator	[redacted]	31/08/2024
Professor Zoe Paskins	Chief Investigator	[redacted]	31/08/2024

The undersigned have approved the content of the analysis plan:

Name	Role	Signature	Date
[redacted]	TSC Chair (on behalf of the TSC)		
[redacted]	DMC Chair (on behalf of the DMC)		

*Signatures may be in the form of an e-mail of endorsement from the TSC/DMC chair that will be printed and stored in the study master file

Data from the iFraP study will be analysed by Elaine Nicholls. To ensure that the study findings are reproducible, analysis of the primary outcome will be replicated by a second, independent statistician, at Keele Clinical Trials Unit but not part of the study team.

Contents

1 Introduction

1.1 Background and rationale

In the UK, three million people are estimated to have osteoporosis, contributing to over 500,000 fragility fractures (fractures resulting from low trauma) per year, costing an estimated £4.4 billion per annum. Fragility fractures can be devastating, sometimes resulting in loss of independence and mortality. Evidence-based treatments, such as bisphosphonates, are recommended by the National Institute for Health and Care Excellence (NICE) for patients with osteoporosis and/or a high fracture risk. They are inexpensive, cost-effective, readily available and reduce fracture risk by 20-70% (depending on fracture site). Despite this, osteoporosis is under-treated with a large 'treatment gap', characterised by the proportion of high-risk patients in whom clinical guidelines would recommend drug treatment who remain untreated. Up to 80% of patients who experience a fragility fracture do not receive medication in the year following fracture. In people who are offered medicine, 25% of people decline it (non-initiation), and among those who do start bisphosphonates, few persist for long enough for it to be effective, with adherence estimated at 16-60% at one year. Closing this treatment gap may prevent at least 20,000 hip fractures annually in the UK.

Although patients ultimately decide whether to start and continue taking medication, this decision making is influenced by the clinician-patient interaction. Effective communication that enables patients to understand complex medical terms and concepts in lay terms and facilitates participation in the consultation may increase patients' commitment to medication. With this in mind, our team developed a package of resources, including a new theoretically-informed computerised decision support tool (CDST), and clinician training programme, in line with guidelines for developing and evaluating complex interventions (iFraP). We hypothesised that the iFraP intervention would facilitate shared decision making, improving patient ease in decision making about osteoporosis medicines (by increasing the extent that the patient was informed and involved in the consultation), increasing informed treatment initiation and reduce treatment discontinuation.

1.2 Objectives

The overall aim of the iFraP study is to examine the experience of care, effectiveness, within-trial cost-effectiveness, and value of information of the iFraP intervention compared with usual FLS practice.

Primary objectives

1) To determine the effect of the iFraP intervention on patient reported ease in decision making about osteoporosis medicines.
2) To determine the cost-effectiveness of iFraP intervention compared to usual Fracture Liaison Services; and the value of acquiring additional information (i.e. value of information (VoI)) on iFraP's cost-effectiveness.

Secondary objectives

1) To determine the effect of the iFraP intervention on a range of patient reported outcomes and experience measures including provision of person-centred care, satisfaction with information, and illness and treatment beliefs.
2) To determine the clinical effectiveness of the iFraP intervention on adherence including treatment initiation and discontinuation rates.

3) To determine the acceptability of iFraP for patients and clinicians, and explore the mechanisms and processes underlying observed effects.
4) To determine clinician adherence to iFraP and clinical guidelines, including the fidelity of the delivered iFraP intervention, and to explore the mechanisms and processes underlying observed effects.
5) To determine barriers and enablers to implementation of iFraP.

1.3 Context

The iFraP trial includes clinical and cost-effectiveness analyses along with a process evaluation. Protocols for the health economics analysis and process evaluation will be published in NIHR Open, hence, this analysis plan relates only to primary objective one, and secondary objectives one and two.

1.4 Estimands for the primary outcome at the primary end-point

Table 1.1: Estimands for the primary outcome at the primary endpoint based on the ICH E9 statistical principles for clinical trials.

Attribute	
Treatment	**Usual Fracture liaison service (FLS) practice** compared to an **FLS consultation intervention that aims to facilitate shared decision making about osteoporosis drug treatment (iFraP),** in the context of usual FLS service provision. Usual FLS Service provision varies across FLSs, with services ranging from operating a 'one-stop shop' model of care, meaning that, if appropriate, patients have a bone density scan (DXA), nurse assessment, drug treatment recommendation, and blood tests as part of one consultation. Other FLS models may not complete all components for all patients (for example, not all patients receive a DXA scan), or may split these components across multiple appointments, supported by different communication modalities (remote, face-to-face, letter). Further details of the interventions are described in the study protocol (Bullock et al. 2024) but in brief, iFraP includes a decision support tool, patient information resources, delivered in a consultation by a clinician who received training in enhanced consultation skills and how to use the tool.
Population	**Analysis population 1** Defined by the inclusion and exclusion criteria of the trial and by excluding patients deemed subsequently ineligible for the study **Inclusion Criteria:** • Adult patients aged ≥50 years eligible for FLS consultation based on having a previous fragility fracture(s)

	• Adult patients able to participate in an FLS appointment (face-to-face or remote consultation) with a participating NHS hospital or associated FLS **Exclusion Criteria:** • Patients who are unable to give full informed consent or unable to comply with study procedures • Patients with a friend or relative in the study (identified through self-report) **Subsequently ineligible** • Patients deemed to be "subsequently ineligible" to include patients that did not attend the FLS appointment or patients that attended the FLS appointment but who had already received an FLS appointment for the same fracture incident **Analysis population 2** This analysis population is a subset of analysis population 1, and is defined as those patients self-reporting on the 2-week questionnaire that they were recommended to start, continue or change their osteoporosis medication (in Section C question 1)
Outcome	Decisional Conflict Scale (DCS) at 2-week follow-up
Population-level summary	Mean Difference (covariate adjusted)
Intercurrent events	Handled using the strategies in section 2.4. A Principal Stratum approach is used to define the analysis populations. Intercurrent events are largely handled using a Treatment Policy approach, except for the event of death, which is treated using a "while alive" strategy

1.5 Trial design

The iFraP trial is a pragmatic, individually randomised, parallel group 2-arm superiority trial in patients referred to UK FLS (with nested process evaluation and health economics evaluation). The trial is designed to test for superiority of iFraP over usual care. The null hypothesis is therefore of no treatment difference. The alternative hypothesis is that there is a difference in outcomes between treatment arms. All statistical tests are 2-sided, with a 5% significance level, hence results are presented using 95% confidence intervals.

1.6 Randomisation

Full details of the randomisation process are given in the study protocol (Bullock et al. 2024). Briefly, patients are randomised with a 1:1 treatment allocation to iFraP or usual care using random permuted blocks stratified by FLS site (block sizes used in the randomisation will be reported when the trial is

complete). The randomisation schedule is determined prior to the trial commencing and according to Keele University's standard operating procedures (SOPs). The randomisation schedule is stored in a password protected file only accessible to the database developer.

1.7 Sample size

This study is powered to detect a between group effect size of at least 0.4 in the primary outcome at 2-week follow-up, with 2-tailed 5% significance and 80% power. With an estimated standard deviation of 15 (Cranney et al. 2002; Kunneman et al. 2020), this translates to minimum clinically important difference (MCID) of 6 points on the DCS (scale range 0 – 100) – a difference considered by the study team to be meaningful and one that produces an effect size in the range of meaningful effect sizes recommended by the authors of the tool (O'Connor 1993).

To achieve an effect size of 0.4 between the study arms, we planned to randomise 380 patients. This recruitment target assumes that approximately 27% of patients will not receive a treatment recommendation (hence for whom the primary outcome is not relevant), for 20% loss to follow-up in the primary outcome at 2-weeks and for 10% loss due to DNAs (patients who 'Do Not Attend); our target at 2-week follow-up is therefore 200 i.e. 100 per arm.

1.8 Framework

All tests of clinical effectiveness will be based on a hypothesis of superiority. No tests of equivalence or non-inferiority will be performed.

1.9 Interim analyses and stopping rules

No interim analysis of treatment effectiveness is planned before the end of the trial.

1.10 Timing of analysis

Treatment effectiveness analyses will be conducted blind to treatment arm (i.e. the statistician will have knowledge of patient arm but will have no knowledge of which arm is iFraP, and which arm is usual care). Analysis will be conducted after all study data has been entered onto the database and after all data queries relating to the effectiveness analysis have been resolved.

1.11 Timing of outcome assessments

Primary and secondary outcomes for the clinical effectiveness analysis are listed in Table 1.2. Additional secondary outcomes include lifestyle-related outcomes, self-perceived fracture risk, worry about falls and fractures, and specific osteoporosis values; listed separately here as the method to analyse them differs from those outcomes listed in Table 1.2 (see sections 5.3 to 5.6 for further details).

Table 1.2: Outcome measures to assess clinical effectiveness.

	Baseline	2-weeks	3-months
Primary outcome (analysis population 2)			
Decisional conflict		x	
Secondary outcomes for analysis population 1			
Patient-Professional Interaction Questionnaire (PPIQ)		x	
Satisfaction with amount of verbal information		x	
Satisfaction with consultation experience		x	
Satisfaction with the amount of written information			x
Modified brief illness perceptions (timeline, consequences, personal control, treatment control, emotional representations, understanding, illness coherence, concern, causal/identity)	x	x	x
Secondary outcomes for analysis population 2			
Beliefs about medicines (BMQ: specific subscale)			x
Satisfaction with medicines information (SIMS)		x	
Self-reported: medicine initiation or intention to initiate		x	x
Self-reported: adherence (MARS-5)			x
Self-reported: persistence or discontinuation with medicine			x
Medical record review: medication initiation			x
Medical record review: medication persistence (discontinuation)			x

Footnote: References for the validated outcomes are given in the study protocol

2 Statistical Principles

2.1 Confidence intervals and p-values

All statistical tests will be 2-sided and tested with 5% significance and presented with 95% confidence intervals. We do not plan to adjust our significance level to account for multiple testing as our outcomes and research hypotheses are pre-planned and specified *a priori*.

2.2 Adherence

Medication adherence is an outcome measure in this trial and will be assessed using self-reported data from the 3-month questionnaire, along with data from a review of (consenting) trial patients' medical records. Further details on how medication adherence is defined is given in section 4.

2.3 Protocol deviations

Patients in the usual care arm will be defined as “treated per protocol” if they did not receive the iFraP intervention

Patients in the iFraP intervention will be defined as “treated per protocol” if it has been recorded on the treatment case report form that the iFraP tool has been used (either partially or fully) in the consultation.

2.4 Analysis populations

We have defined our analysis populations using the ICH E9 (R1) addendum on estimands (Clark et al. 2022). The analysis strategies are described in Table 2.1 and how they apply to our analysis populations is given in Table 2.2.

Table 2.1: Analysis strategies

Strategy	Implementation in the data
Treatment policy	The value for the variable of interest will be used in the analysis regardless of whether the intercurrent event occurs
While-alive	Any data collected after a patient has died will remain as missing data in the analysis and not imputed using multiple imputation
Principal stratum	Patients meeting the definition for the “principal stratum” will be included in the dataset

Table 2.2: Analysis populations definitions.

Post-randomisation intercurrent events	Analysis population 1	Analysis population 2
Subsequently ineligible (as defined in Table 1.1)	Principal Stratum	Principal Stratum
Recommendation given at the FLS appointment to take osteoporosis medication	Treatment policy	Principal Stratum
Did not take the osteoporosis medication that was recommended for them	Treatment policy	Treatment policy
Use of other medication and supplements	Treatment policy	Treatment policy
Protocol deviations that impact primary and secondary outcome data collection	Treatment policy	Treatment policy
Adverse events	Treatment policy	Treatment policy
Death[α]	While alive	While alive

[α] We chose a 'while alive' strategy for death to avoid applying the unrealistic assumption of an immortal cohort (Wen L et al. 2017)

3 Trial Population

A CONSORT flow diagram will be produced (Figure 13.1) to describe the number of patients recruited and followed up, along with reasons for ineligibility or withdrawal (if given) (Schulz et al. 2010).

Patients' baseline characteristics will be described for analysis populations 1 and 2 using the variables outlined in Table 13.1 and Table 14.1. The data in Table 13.1 will also be stratified by treatment arm, and whether patients were lost to follow-up at 2-weeks and 3-months (defined by whether the patient returned the questionnaire at this time-point), to explore balance in the randomisation process and if there is any evidence of selective loss to follow-up.

The key characteristics of age, sex at birth and the index of multiple deprivation (where collected) of patients included at each recruitment stage will be described (Table 14.2).

All analyses in this section will use descriptive statistics only (i.e. numbers and percentages for categorical data, means and standard deviations for normally distributed continuous data and median and inter-quartile range for skewed continuous data) hence no statistical tests will be used to compare statistical significance of patient characteristics across subgroups of the data.

4 Outcome definitions

4.1 Derivation rules

Validated trial outcome measures will be scored using the instructions from the authors of the tool. Prior to outcome scoring, we will apply (where needed), principles in our internal Standard Operating Procedure (SOP) 16 – Data Analysis – Version 5.0, to process the data e.g. to handle responses where two answers have been given to a single question. We will document the outcome of this decision-making process in the analysis syntax file to ensure that the trial results can be replicated. A description of the scoring of all outcome measures is given in Table 4.1.

Table 4.1: Description of the derivation of study outcome measures and other derived measures used in the trial analysis.

Outcome measure	Scoring rule	Missing data considerations	Score interpretation	Scoring reference website (if applicable)
Outcome measures calculated in analysis population 1 (i.e. all patients)				
Patient-Professional Interaction Questionnaire (PPIQ)	Coded using the instructions in the scoring reference.	No guide on how to handle missing data is provided in the tool. Score calculated if 12 or more of the 16 items are present[α]	Range: 16 – 80 Higher score: greater professional interaction experienced	https://www.sciencedirect.com/science/article/pii/S0738399118304725
Satisfaction with amount of verbal information	A subset of six items from the satisfaction with cancer information profile (SCIP) tool was used. Items coded from 1 = Very dissatisfied to 5 = Very satisfied and summed into a total score.	Missing data scoring rule provided in the tool: score calculated if 4 or more of the 6 items are present	Range: 6 – 30 Higher score: greater satisfaction	https://onlinelibrary.wiley.com/doi/10.1002/hed.20450
Satisfaction with amount of written information	A subset of three items from the satisfaction with cancer information profile (SCIP) tool was used. Items coded from 1 = Very dissatisfied to 5 = Very satisfied and summed into a total score.	All items to be present for a score to be calculated[α]	Range: 3 – 15 Higher score: greater satisfaction	https://onlinelibrary.wiley.com/doi/10.1002/hed.20450
Outcome measures calculated in analysis population 2 (i.e. in patients given a drug recommendation)				
Decisional conflict scale (DCS)	Coded using the instructions in the scoring reference.	No guide on how to handle missing data is provided in the tool. Score calculated if 12 or more of the 16 items are present[α].	Range: 0 – 100 Higher score: greater decisional conflict	https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf
Beliefs and Medicines questionnaire (BMQ)	Scores calculated separately for the general and specific subscales (8 and 11 items respectively). Items summed to a total score for each subscale	No guide on how to handle missing data is provided in the tool. BMQ-general calculated if 7 or more items are present[α] BMQ-specific calculated if 9 items or more are present[α]	BMQ-general: Range: 8 - 40 BMQ-specific: Range: 11 – 55 Higher score: stronger belief in the subscale concept	https://www.tandfonline.com/doi/epdf/10.1080/08870449908407311?needAccess=true

The Satisfaction with Information about Medicines Scale (SIMS)	Scale calculated using the instructions in the scoring reference	All 17 items need to be present for a score to be calculated (within-person mean value estimation is not possible for this measure given that the score is a count across items, rather than a total score)	Range: 0 – 17 Higher score: greater satisfaction	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1743429/pdf/v010p00135.pdf
Self-reported: medicine initiation or intention to initiate	Medication initiation defined if patient report "yes" they are currently taking medication OR "yes" that they intend to take the recommended medicine (2-week questionnaire Section C. Q1a and Q1b; 3-month questionnaire Section E Q1 and Q1a)	Defined as missing if either of the two questions required for the definition are missing	Binary outcome (yes/no)	Not applicable
Self-reported: medicine persistence/discontinuation	Defined as "yes" discontinued if the patient reports that they have stopped using all types of osteoporosis medication	Defined as missing if the 3-month questionnaire is missing	3-level categorical outcome: (did not start taking medication; discontinued medication; continued medication)	Not applicable
Self-reported medication adherence (MARS-5)	Scale calculated using the instructions in the scoring reference	No guide on how to handle missing data is provided in the tool. Score calculated if 4 or more items are present [α]	Range: 5 - 25 Higher score: greater medication adherence	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7319010/
Medical record review: medication initiation	Medication initiation indicated as "yes" if use of osteoporosis medication was evident in the medical records	Calculated only for patients with medical record data available for the full (individual-specific) 3-month follow-up period (i.e. the patient did not withdraw consent or die during the study)	Binary outcome (yes/no)	Not applicable

Medical record review: medication discontinuation	Medication discontinuation indicated as "yes" if the last date of prescription is ≥ 6 weeks prior to the patient's 3-month follow-up OR medication discontinuation is recorded in the patient medical record	Calculated only for patients with medical record data available for the full (individual-specific) 3-month follow-up period (i.e. the patient did not withdraw consent or die during the during the study)	3-level categorical outcome: (did not start taking medication; continued medication; discontinued medication)	Not applicable
Descriptive variables				
Age	Age will be defined using the variable "age" as stored in the patient identification module in REDCAP.	We will consider information on date or birth at subsequent time-points to calculate age if there is missing data for the age variable in REDCAP	Range 50 years and over	Not applicable
Sex at birth	Sex at birth will be defined using the variable "sex at birth" as stored in the patient identification module in REDCAP.	We will consider information on sex at birth on the baseline questionnaire if there is missing data for sex at birth in REDCAP	Male/ Female	Not applicable
Index of multiple deprivation (IMD) 2019.	Derived from patient postcode data	All responses used for analysis	Range 1 – 32844 Lower score most deprived Also categorised into quintiles of deprivation 1: IMD 1 to 6568 2: IMD 6569 to 13137 3: IMD 13138 to 19706 4: IMD 19707 to 26275 5: IMD 26276 to 32844	**Research report for 2019 coding:** https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833947/IoD2019_Research_Report.pdf **Scoring calculator (2019 version)**

			This corresponds with combining deciles of IMD into pairs i.e. deciles (1,2); (3,4); (5,6); (7,8); (9,10).	https://imd-by-postcode.opendatacommunities.org/imd/2019
Body mass index	Calculated from weight and height on the baseline questionnaire as (Weight in kgs)/(height in meters)^2	Complete data for weight and height required	kg/m^2 Also categorised as underweight/normal weight/overweight/obese according to the cut-points in the scoring reference	Weir et al. 2022

Footnote: α = Where guidance has been given in the published tool as to how missing data should be handled this has been followed, otherwise guidance given in SOP 16 version 5.0 (Data analysis) has been used to determine the maximum number of missing items to allow in the score calculation. The rules from SOP16 suggest that if a tool contains less than 5 items, all items need to be present for a score to be calculated, between 5 and 10 items, 80% of items need to be present, between 11 and 20 items, 75% of items need to be present, and over 20 items 70% of items need to be present. Where enough items are present for a score to be calculated, the missing values are estimated using the within-person mean value of the items completed

5 Analysis methods (1): clinical effectiveness analysis of Usual care versus iFraP

5.1 Primary analysis

5.1.1 Statistical model

The primary outcome (the total DCS score) will be analysed using Analysis of covariance (ANCOVA) at the 2-week primary endpoint. The model will include a binary term for treatment arm and will be adjusted for the baseline covariates listed in section 5.1.2. The adjusted ANCOVA model will form the primary analysis and will be run on analysis population 2, after multiple imputation of missing data has been applied, and after model assumptions have been checked. Model results will be presented as mean differences and 95% confidence intervals and will be based on estimates combined across the multiply imputed datasets using Rubin's rules (Table 13.2). Descriptive data (e.g. means and standard deviations) will also be reported for the DCS by treatment arm.

5.1.2 Adjusting covariates

Adjusting covariates will include FLS site (as this is a stratification variable used in the generation of the randomisation schedule), along with age (years), sex at birth (male, female), and the index of multiple deprivation.

5.1.3 Checking model assumptions

Model assumptions will be checked just prior to database lock using a model that does not include a term for treatment. This is to ensure that if the model assumptions do not hold, and a change of model is required, then the choice of model is not influenced by the magnitude of the observed treatment effect (e.g. that the results are not inflated to show the largest treatment effect possible in the data). For example, if the outcome of interest follows a skewed distribution, it may be necessary to transform the data prior to analysis e.g. using a log or square root transformation, or to analyse the data using a more complex generalised linear model.

If the model chosen prior to the inclusion of the treatment term, subsequently does not fit the data well when the treatment term is added, and a different model subsequently becomes more suitable, we will state this in the publication that the chosen model had been determined outside the *a priori* analysis plan.

Model assumptions for the ANCOVA model will be checked prior to imputation of missing data. Descriptive plots e.g. histograms and scatter plots will be used to check:

1. That the model residuals follow a normal distribution
2. That no relationship exists between the model residuals and predicted values
3. That no relationship exists between the model residuals and the independent variable in the model and that any variability in the residuals is consistent across the range of values for the independent variables.
4. That a linear relationship is observed between the dependent variable and each predictor in the model to ensure a linear relationship is observed. We note that we do not necessarily assume that there will be a linear relationship between our model predictors and the DCS score (e.g. the relationship between age and DCS score may not necessarily be a linear one).

If this occurs, we will consider adding a quadratic term to the model for our continuous predictors to account for any non-linearity in the data. We will retain the quadratic term in the model if it is statistically significant and improves model fit.

We will also check if any outliers remain in the data after the *a priori* rules for handling implausible values have been applied. If outliers are found, the data will be checked for accuracy, and if no error in the data is found, it will be included in the analysis to reflect the pragmatic nature of the trial.

5.1.4 Missing data

5.1.4.1 Descriptive statistics

The percentage of missing data will be calculated for the primary and secondary outcomes. The denominator for the calculation will be being either the number of patients in analysis population 1 or 2 as relevant for the outcome of interest (analysis populations applicable to each outcome measure are defined in Table 1.2).

5.1.4.2 Multiple imputation

Multiple imputation will be used to impute missing data for the primary and secondary outcomes in Table 1.2.

The imputation model will include:

1) the primary and secondary outcomes of interest at all time-points where data are collected (Table 1.2)[1]
2) the adjusting variables in the regression model (Section 5.1.2)
3) the treating therapist (to enable a sensitivity analysis to be completed)
4) mode of consultation delivery (face to face or telephone), health literacy and whether the patient had a bone health record (to enable subgroup analyses to be completed).

Outcome measures only relevant to analysis population 2 will remain missing in the imputed datasets for patients that do not meet the criteria to be in analysis population 2. We will fit the imputation model separately for each arm of the trial (usual care and iFraP) to enable treatment interactions to be included in our analysis models (White et al. 2011, Cro et al. 2020). Our aim is to ensure that all variables in our analysis models on imputed data are included in the imputation model (Austin et al. 2021).

The imputation model will be fitted using Multiple Imputation by chained equations (MICE), assume the data are missing at random, and will include X imputed datasets. We will use the guidelines in (White et al. 2011) to determine the value of X, to ensure that the Monte Carlo error (MCE) estimates for all parameter estimates are <= 10% of their respective standard errors, that the MCEs for the test statistics are <=0.1 and that the MCEs for the p-values are <= 0.01 for all models fitted on the imputed data. We chose to use MICE as our imputation method, rather than Multi-Variate Normal Imputation (MVNI), as MICE offers greater flexibility to form imputation models outside any known standard multivariate density function (van Buuren et al. 2007)

[1] We chose to impute the data at the level of the scale score (where applicable) rather than at the individual item level. Differences between the two approaches have been shown to be small for large sample sizes, and, by not imputing at the individual item level, we reduce the risk of convergence issues that are likely when imputing many categorical items (Rombach et al. 2018).

The imputation model will include continuous outcome measures, modelled using predictive mean matching (nearest neighbours = 10 (Morris et al. 2014)); binary outcomes, modelled using logistic regression; nominal variables, modelled as multinomial logistic regression and ordinal outcomes, modelled using ordinal regression. Predictive mean matching will be used for continuous measures as this method is suitable for the imputation of both normally distributed and skewed outcomes and produces imputed values restricted to the range of values that the measure requiring imputation can take (Morris et al. 2014).

The imputation model will be fitted to the data, however, given the complexity of the model, it may arise that the imputation model will breakdown, so it may not be possible for it to be fitted to the data. If this occurs, then the techniques described in section 15.1 will be explored to see how the imputation model can be adapted to ensure it can be fitted to the data. If adaptations need to be made to the imputation model, this will be explained in the results publication for the trial. If a successful imputation model can be developed, analysis models will then be fitted, and Rubin's rules (Rubin and Schenker, 1991, Austin et al. 2021) used to combine the treatment effects and their associated standard errors across the imputed data sets. This will provide a single estimate of the treatment effect for each analysis outcome.

5.1.5 Checking the imputation model

Descriptive graphs (histograms, box plots) and statistics (means, standard deviations, ranges) will be used to check that the imputed data for each variable appear theoretically plausible from what is known about the (clinical) range of the scales in the observed data. We will also produce a boxplot of the primary outcome in the observed data and compare this to the equivalent box plot in each imputed dataset to explore whether the distribution of the primary outcome in each imputed dataset is similar, or otherwise, to the observed data (we plan to do this to increase our understanding of the impact that multiple imputation has on our dataset as if the data are missing not at random then it may not be of concern if the imputed data differ from the observed data).

5.1.6 Sensitivity analyses for the primary analysis: Treatment effect for the DCS estimated after variation between treating FLS clinicians has been accounted for in the ANOVA model.

We plan to run a sensitivity analysis to explore the stability of the treatment effect (iFraP versus usual care) for the DCS outcome measure at two weeks after clustering of patients within FLS clinician is considered[2]. The model for the primary analysis (defined in section 5.1) will be re-run but will be converted into a mixed model framework by inclusion of a random effect term representing the FLS clinician who delivered the intervention. The magnitude of the treatment effect from this model will then be compared to that from the primary study analysis, with results presented as a mean difference between iFraP and usual care with the associated 95% confidence interval (Table 13.2).

For this model, there is the potential that the model will not converge if the treating FLS clinician is completely nested within "Treating Centre". If this is the case then a model dropping the term for "Treating Centre" will also be run, and the results of this latter model reported as an alternative model.

[2] We do not plan to run a sensitivity analysis assuming that missing data are "missing not at random" as we anticipate a high follow-up response rate at 2-weeks so the method used to estimate the missing data will have minimal effect on the trial results presented

In this trial the FLS clinicians are trained to deliver one arm of the trial only. We did not inflate our sample size to account for this, as practically we could not deliver the large sample size that would be required to inflate the sample size to accommodate for this, particularly for an intervention that is considered low risk. Consequently, our analysis adjusting for clustering by FLS clinician remains exploratory.

5.2 Secondary analysis

The description below of the secondary analysis methods applies only to the outcome measures listed in Table 1.2

Table 1.2.

5.2.1 Continuous outcomes: data collected at a single follow-up time-point

Treatment effectiveness for continuous outcome measures, listed in Table 1.2, that are measured at a single follow-up time-point will use the same overall method of analysis as for the primary outcome (Section 5.1), with the ANCOVA model fitted to the analysis population of interest relevant to the outcome modelled e.g. the Patient-Professional Interaction Questionnaire (PPIQ) will be analysed in analysis population 1, whereas, the Beliefs about medicines (BMQ: specific subscale) outcome will be analysed in analysis population 2.

5.2.2 Continuous outcomes: data collected at a multiple follow-up time-points

Mixed models will be used to model continuous outcome measures that are collected at both the 2-week and 3-month follow-up i.e. the subscales of the illness perceptions questionnaire. Separate models will be fitted for each sub-scale score. The outcome (defined in "long format" at 2-weeks and 3-months) will be predicted from: fixed effects terms for treatment (usual care vs iFraP), time (2-weeks, or 3-months), the interaction between treatment and time, baseline for the subscale score of interest, and the adjusting covariates listed in section 5.1.2; and a random effect term for the intercept to reflect the lack of independence in the data (Twisk J et al. 2018).

To generate an estimate of the treatment effect at 2-weeks, time will be coded as 0 = 2-weeks, 1 = 3-months. To generate an estimate of the treatment effect at 3-months, time will be coded as 1 = 2-weeks, 0 = 3-months. The treatment effects will be presented alongside 95% confidence intervals.

A plot will also be produced of each illness perception subscale over time to explore firstly, whether patients in the iFraP arm of the trial show greater changes at 2-weeks towards a positive views of their bone health compared to patients in the usual care; and secondly, whether any positive changes observed at the 2-week follow-up are more likely to be sustained, longer-term, at the 3-month follow, in the iFraP arm compared to usual care.

The mixed models will be fitted using full information maximum likelihood (FIML) estimation if the illness perceptions subscales follow a normal distribution, or FIML with robust standard errors if this assumption is not met (as evaluated by visual inspection of a histogram of the illness perceptions subscales). Separate residual terms will be fitted for each follow-up time-point, however if this model fails to converge, the model will be simplified to assume a common residual across all time-points.

Assumptions for the mixed model will be explored as below, and if not met, this will be reported (Singer 2003):

1. A histogram of model residuals and random intercepts (estimated using empirical Bayes estimation/best unbiased linear predictors (BLUPs)) will be produced to ensure they are normally distributed.
2. Plots of the model residuals and random intercepts against study identification number will be generated to ensure no relationship exists and to identify any specific patients with large residuals or random intercepts (i.e., to check for outliers) (the model will not be re-run excluding outliers as this is a pragmatic trial, but if large, the number of outliers will be reported).
3. Plots of the random intercepts against time-invariant predictors in the model and the residuals by time. No relationship should exist in these plots; they will also be used to check whether the assumption of homogeneity of variance holds for each variable in the model.
4. The covariance between the residuals and the random intercepts in the model will be inspected to ensure it is close to 0 (such covariance is assumed to be 0 in the models as fitted).

5.2.3 Binary outcomes

Binary outcome measures will be modelled using logistic regression and will include a predictor-term for treatment and the adjusting covariates listed in 5.1.2. Results will be presented as numbers, percentages, and odds ratios with 95% confidence intervals.

Model assumptions for the logistic regression models will explore if there is a linear relationship between the log odds and each independent variable in the model. A plot of the residuals will also be used to identify any outliers in the data, but such outliers will remain included in the analysis to reflect the pragmatic nature of the trial.

5.2.4 Ordinal outcomes

Ordinal outcomes will be modelled using ordinal regression. Model predictors will include treatment arm and the adjusting covariates listed in 5.1.2. Prior to analysis, category response frequencies will be explored, and appropriate category merging will be conducted for any response category with a response frequency <10% with the aim to encourage stable model estimates to be derived. Model results for the treatment effect estimate will be presented as odds ratios with associated 95% confidence intervals.

It will be explored whether the proportional odds assumption is satisfied for the treatment variable in the ordinal model by exploring whether model fit significantly improves if a separate effect for treatment is estimated for each cut-point within the dependent variable (i.e. if a partial proportional odds model is fitted to the data). If model fit significantly improves by relaxing this assumption (as tested using a likelihood ratio test), results from the partial proportional odds model will also be reported. If the POM or partial POM models fail to converge (which could be a potential problem if the number of categories in the model is large), then multinomial or logistic regression will be considered as an alternative analysis option.

Results from the secondary analysis described in sections 5.2 will be presented using outline Table 13.2 and Table 13.3.

5.3 Secondary outcome: lifestyle outcome measures

We hypothesise that the iFraP intervention will encourage patients to make lifestyle changes where this is required, namely:

- reduce alcohol consumption if the patient is drinking heavily
- giving up smoking if the patient is currently smoking,
- increase weight if the patient is underweight
- increase levels of physical activity if the patient is sedentary.

As these outcome measures may only apply to a subset of the analysis population (e.g. giving up smoking is only an outcome for patients who smoke) they require a different analysis approach to other outcomes in the trial. The analysis approach taken will be descriptive as there is the potential that some of the analyses will be based on a small sample-sizes so may not be reliable.

We will describe the following proportions, based on analysis population 1 and stratified by treatment arm as outlined in Table 14.3:

1) The number patients that have reduced their alcohol use as a proportion of patients in analysis population 1 that report drinking daily or on most days. Alcohol use is measured on a 5-point Likert scale (*frequency: daily or most days, once or twice a week, once or twice a month, once or twice a year, never*), so reduction in alcohol use will be defined by a reduction, from baseline, of one point or more on the 5-point Likert scale when asked at the 3-month follow-up
2) The number of patients that are not currently smoking at 3-month follow-up as a proportion of those who were currently smoking at baseline
3) The number of patients that are underweight at 3-month follow-up as a proportion of those that were underweight at baseline. We will define being underweight as having a body mass index less than 18.5kg/m^2
4) The proportion of patients that report "yes" they have increased their physical activity in analysis population 1. We will also report the reasons given for this change (i.e. for weight loss, to improve bone health, or for other health reasons)

5.4 Secondary outcome: self-perceived fracture risk

We aim to explore whether patients' perception of their fracture risk changes following the FLS consultation. We will address this using descriptive data from the question: *Compared to other people of the same sex and age as you, do you consider your chance of breaking a bone is: (1) Much lower, (2) A little lower, (3) About the same, (4) A little higher, (5) Much higher.* We will report the percentage of patients in analysis population 1 whose perception of risk changed between baseline and the 2-week follow-up by treatment arm, and will sub-divide this into those who thought their risk increased i.e. that moved to a response option with a higher number; and those who thought their risk decreased i.e. that moved to a response option with a lower number; and those who thought their risk remained the same (Table 14.4).

We will also explore whether patient's post FLS consultation perception of risk is more in line with their predicted fracture risk (as calculated by their FRAX score). Patient FRAX scores (entered by their treating clinician and calculated using the FRAX algorithm (https://www.sheffield.ac.uk/FRAX/)) will be categorised into three groups (low, middle, high risk) based on standard cut-offs (NOGG assessment and treatment thresholds (Gregson et al, 2022). This variable will then be compared to a patients perceived fracture risk using the variable as described in the paragraph above (categorised as lower risk (categories 1 and 2), the same risk (category 3), higher risk (categories 4 and 5) and stratified by treatment arm (Table 14.5)

5.5 Secondary outcome: worry about falls and fractures

An unintended consequence of the iFraP intervention may be that patients become more concerned (worried) about having a fall and breaking a bone after the consultation as they are now more informed about the risks. We will test this hypothesis descriptively by calculating, for each outcome separately (worry about falls and worry about breaking a bone) the number and proportion of patients whose worry increases by 1, 2, 3, 4 or 5 points between the question asked at baseline and question asked at the 2-week follow-up (the question about worry is measured on a scale of 1 to 6 with 1 = Not worried at all to 6 = Very worried). Similarly, the percentage whose worry lessens will also be calculated for context in the data. These proportions will be compared between the treatment arms descriptively to evaluate whether the unintended consequence of increased worry about falls and breaking a bone in the iFraP arm is evidenced in the data. This analysis will be completed in analysis population 2, i.e. in those patients that have been given a drug recommendation as it is in this population that an increase in worry about falls and fractures is most likely to be observed (Table 14.6).

5.6 Secondary outcome: specific osteoporosis values

Osteoporosis values, relating to perceptions of osteoporosis medicine benefits and harms will be described using numbers and percentages. The results will be presented at the 2-week time-point, for analysis population 2 and stratified by treatment arm (Table 14.7).

5.7 Supplementary/exploratory analysis

5.7.1 Treatment effect (iFraP vs usual care) for the primary outcome (DCS) at the 2-week follow-up within key patient subgroups of interest

Exploratory subgroup analyses will be performed for the primary outcome (DCS) at the 2-week follow-up to test whether the magnitude of the treatment effect depends on the subgrouping variable of interest. We acknowledge the exploratory nature of these subgroup analyses given the relatively small number of patients in each treatment arm (approximately 100 patients) that have completed the DCS that may result in there being insufficient power to statistically detect whether the interactions exist in the data.

That said, our sub-grouping variables of interest are:

1. **Mode of consultation delivery**: face-to-face or telephone (*we anticipate that the treatment effect may be larger for face-to-face consultations rather than those delivered over the telephone as those delivered over the telephone were not able to see the visual aspects of the iFraP tool).*
2. **Bone health record:** received a bone health record received (yes/no) (*we anticipate that the treatment effect may be larger for those that received the bone health record than those that did not)*
3. **Health literacy**: never vs some difficulty *(we anticipate that the treatment effect may be larger for patients reporting no problems with health literacy)*. We are specifically interested in this subgroup analysis as it will allow us to compare our findings to an external study on decision aids and medication persistence in FLS (Cornelissen et al. 2021)

4. **Sex at birth:** Male or female. We have included this subgroup analysis based on journal recommendations (e.g. by the British Medical Journal (BMJ) to present data stratified by sex, but do not have any *a priori* expectations around the direction of this effect

The primary analysis model, described in section 5.1, will be re-fitted to the data but will include the subgrouping variable and an interaction term between treatment and each subgrouping variable of interest (separate models for each subgrouping variable). Model results will be presented as parameter estimates for the interaction terms, along with associated 95% confidence intervals (Table 14.8),

Prior to analysis, we anticipate two potential problems with the models described above that we will explore:

1) For the model where mode of consultation delivery is explored, we know that the mode of consultation delivery is directly related to the FLS site i.e. at one site all consultations were delivered face to face, whereas at another site all consultations were delivered on the telephone. For this reason, if the model may not run with both site and mode of consultation in it. If this is the case, then site will be dropped from the model and only the mode of consultation delivery explored.
2) For the health literacy model, we anticipate that health literacy could be related to the index of multiple deprivation. Therefore, we will run a model with and without the index of multiple deprivation in it, to explore the impact that this could have on the study findings.

5.7.2 Mode of treatment delivery

A secondary question to explore is whether the model of treatment delivery has an impact on the absolute magnitude of the trial outcomes, rather than the magnitude of the treatment effect in the trial. This will be important to understand as it will help inform how the iFraP intervention could be delivered in practice after the trial.

Given that mode of delivery directly relates to the FLS site, we will explore this research question by examining the regression coefficients for the Site variable in the regression model that was used to estimate the treatment effect for the DCS outcome at the 2-week follow-up. We will hypothesise that the regression coefficient for the site that used face to face consultations will be larger than for the sites that used telephone consultations as an alternative.

6 Safety

6.1 Serious and unexpected adverse events

This trial is of a low-risk intervention so serious and unexpected adverse events are likely to be rare.

We will follow the safety reporting procedures as specified in the study protocol. The number and percentage of patients experiencing a serious adverse event (SAE) that was judged to be related to the treatment intervention will be reported, both overall and by treatment arm (the denominator for the percentages is the number of patients that attended an FLS appointment). Details of each event will be described in text or table as appropriate. We do not anticipate that many patients will experience multiple SAEs, but if they do, then the number of SAEs that each person experienced will be reported as a percentage of patients experiencing one or more SAEs. No formal statistical testing will be used

to test whether the number of adverse events differs between treatment arms; hence the percentage of adverse events will be evaluated descriptively and assessed for clinical significance.

6.2 Medicine side effects

A list of closed-form osteoporosis medication side-effects are included in the 3-month questionnaire, that seek information on whether the side effect occurred, and whether the patient believes that the side effect is caused by osteoporosis or their osteoporosis medicine. Patients reporting that they have taken osteoporosis medicines since their FLS appointment (as reported on their 3-month questionnaire) will form the sample for the analysis. Medication side effects will be described for each treatment arm in the trial as shown in outline Table 14.9.

7 Protocol deviations

A list of all protocol deviations will be produced by arm, and each will be judged according to whether they are a major or minor deviation, and whether the deviation is likely to affect responses given to the primary outcome. Only protocol deviations classified as major deviations will be reported in the publication arising from the study. No formal statistical testing will be undertaken to test whether the proportion of patients treated per protocol differs by treatment arm

8 Further research questions of interest

This analysis plan does not include exploratory analyses that are part of the process evaluation to explore mechanisms of action. For example, further work, outside the main trial, will explore factors that predict treatment self-report initiation at the 3-month follow-up in those patients given a drug recommendation, with logistic regression used to predict treatment initiation at 3-months from a list of candidate predictors to include age, sex, fracture risk, self-reported receipt of osteoporosis diagnosis, consultation length, consultation modality (face-to-face vs telephone), receipt of a DXA scan, level of health literacy and socioeconomic status. In addition, we will use linear regression to explore whether factors (such as age, sex, health literacy and socioeconomic status) are associated with the primary outcome of interest (the DCS). We will also potentially explore how self-reported osteoporosis values compare with those recorded in the consultation by the clinician on the bone health record.

9 Software

Analysis in this analysis plan will be generated using STATA software and will use the most up-to-date version of the software available for analysis. The software version number will be reported in any published papers arising from the trial.

10 Data management plan

Trial data collection followed Data Management Plan (DMP) version 1.0 – 24th May 2023.

11 References

Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: A tutorial on multiple imputation. Can J Cardiol. 2021;37(9):1322-1331. doi: S0828-282X(20)31111-9 [pii].

Bullock L, Nicholls E, Cherrington A, et al. A person-centred consultation intervention to improve shared decision-making about, and uptake of, osteoporosis medicines (iFraP): A pragmatic, parallel-group, individual randomised controlled trial protocol [version 1; peer review: 2 approved]. *NIHR Open Research*. 2024;4(14). https://openresearch.nihr.ac.uk/articles/4-14/v1. doi: 10.3310/nihropenres.13571.1.

Clark TP, Kahan BC, Phillips A, White I, Carpenter JR. Estimands: Bringing clarity and focus to research questions in clinical trials. BMJ Open. 2022;12(1):e052953. http://bmjopen.bmj.com/content/12/1/e052953.abstract. doi: 10.1136/bmjopen-2021-052953.

Cornelissen, D., Boonen, A., Evers, S. *et al.* Improvement of osteoporosis Care Organized by Nurses: ICON study - Protocol of a quasi-experimental study to assess the (cost)-effectiveness of combining a decision aid with motivational interviewing for improving medication persistence in patients with a recent fracture being treated at the fracture liaison service. *BMC Musculoskelet Disord* **22**, 913 (2021). https://doi.org/10.1186/s12891-021-04743-2

Cranney A, O'Connor AM, Jacobsen MJ, et al. Development and pilot testing of a decision aid for postmenopausal women with osteoporosis. *Patient Educ Couns*. 2002;47(3):245-255. doi: 10.1016/s0738-3991(01)00218-x.

Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Statistics in Medicine. 2020;39(21):2815-2842. https://doi.org/10.1002/sim.8569. doi: 10.1002/sim.8569.

Gregson CL, Armstrong DJ, Bowden J, Cooper C, Edwards J, Gittoes NJL, Harvey N, Kanis J, Leyland S, Low R, McCloskey E, Moss K, Parker J, Paskins Z, Poole K, Reid DM, Stone M, Thomson J, Vine N, Compston J. UK clinical guideline for the prevention and treatment of osteoporosis. Arch Osteoporos. 2022 Apr 5;17(1):58. doi: 10.1007/s11657-022-01061-5. Erratum in: Arch Osteoporos. 2022 May 19;17(1):80. doi: 10.1007/s11657-022-01115-8. PMID: 35378630; PMCID: PMC8979902.

Kleinke K. Multiple imputation by predictive mean matching when sample size is small. Methodology. 2018;14(1):3-15. https://doi.org/10.1027/1614-2241/a000141. doi: 10.1027/1614-2241/a000141.

Kunneman M, Branda ME, Hargraves IG, et al. Assessment of shared decision-making for stroke prevention in patients with atrial fibrillation: A randomized clinical trial. *JAMA Intern Med*. 2020;180(9):1215-1224. doi: 10.1001/jamainternmed.2020.2908.

Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Medical Research Methodology. 2014;14(1):75. https://doi.org/10.1186/1471-2288-14-75. doi: 10.1186/1471-2288-14-75.

Nguyen CD, Carlin JB, Lee KJ. Practical strategies for handling breakdown of multiple imputation procedures. Emerging Themes in Epidemiology. 2021;18(1):5. https://doi.org/10.1186/s12982-021-00095-3. doi: 10.1186/s12982-021-00095-3.

O'Connor AM: User Manual - Decisional Conflict Scale. Ottawa Hospital Research Institute, 1993; accessed 27 February 2020.

Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. Multiple imputation for patient reported outcome measures in randomised controlled trials: Advantages and disadvantages of imputing at the item, subscale or composite score level. BMC Med Res Methodol. 2018;18(1):87-6. doi: 10.1186/s12874-018-0542-6 [doi].

Rubin D, Schenker N. Multiple imputation in health-are databases: An overview and some applications. Stat Med 1991; 10:585–98.

Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332. http://www.bmj.com/content/340/bmj.c332.abstract. doi: 10.1136/bmj.c332.

Singer, J., & Willet, J. (2003). Applied longitudinal data analysis (First ed.). USA: Oxford University Press.

StataCorp. 2022. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.

Twisk J; Bosman L; Hoekstra T; Rijnhart J; Welten M and Heymans M. Different ways to estimate treatment effects in randomised controlled trials. Contemporary Clinical Trials Communications. 2018;10:80-85. https://www.sciencedirect.com/science/article/pii/S2451865417301849. doi: 10.1016/j.conctc.2018.03.008.

van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007;16(3):219-242. https://doi.org/10.1177/0962280206074463. doi: 10.1177/0962280206074463.

Weir CB, Jan A. BMI classification percentile and cut off points. In: StatPearls. Treasure Island (FL): StatPearls Publishing LLC; 2022. NBK541070 [bookaccession].

Wen L, Terrera GM, Seaman SR. Methods for handling longitudinal outcome processes truncated by dropout and death. Biostatistics. 2017;19(4):407-425. https://doi.org/10.1093/biostatistics/kxx045. Accessed 4/22/2023. doi: 10.1093/biostatistics/kxx045.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-399. doi: 10.1002/sim.4067 [doi].

12 Acknowledgments

We acknowledge that parts of this analysis plan have been taken and adapted from the analysis plan that was written for the PROP-OA trial (https://www.isrctn.com/ISRCTN28555470).

13 Outline tables and figures (for the main body of the results paper)

Figure 13.1: CONSORT flow diagram

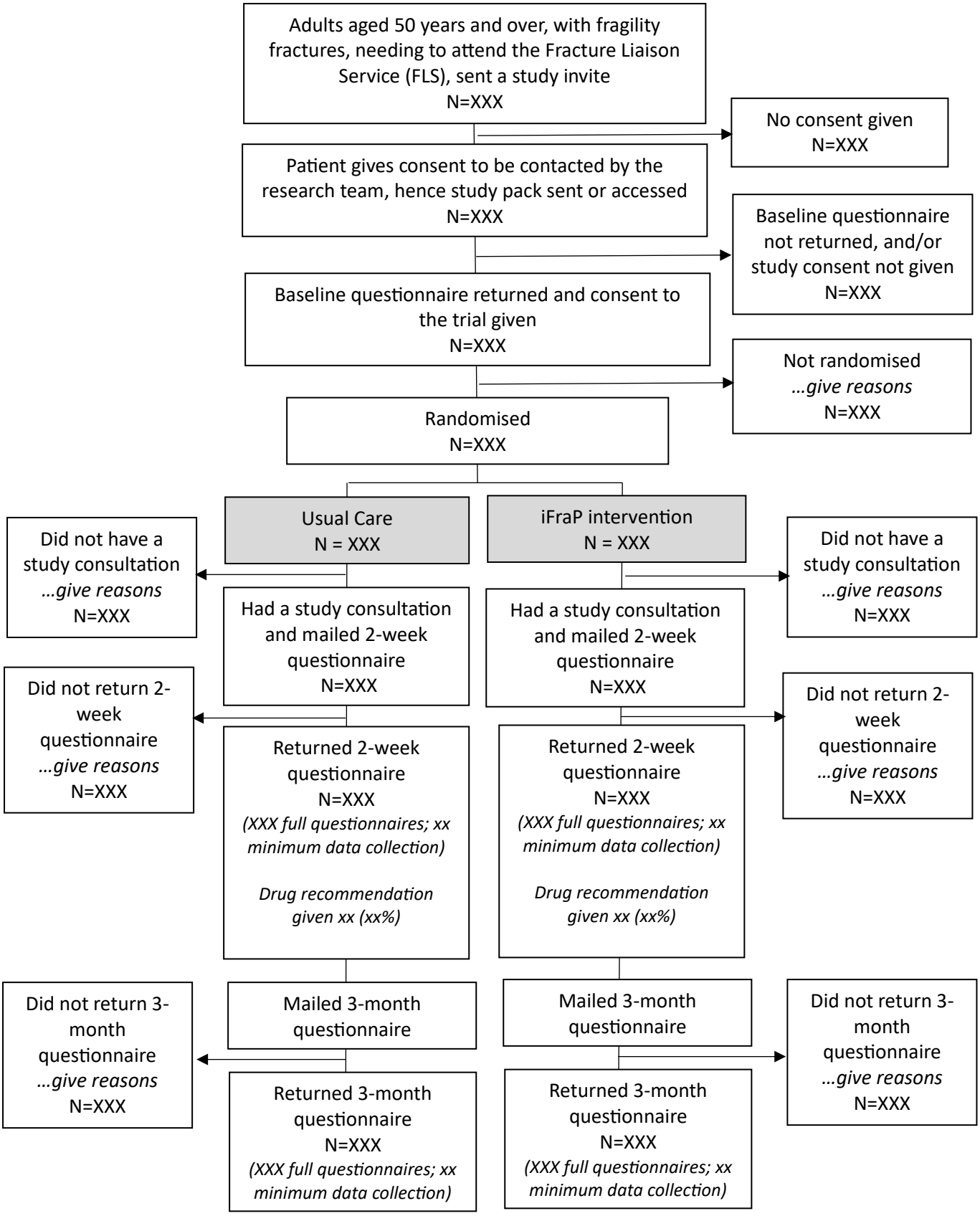

Table 13.1: Baseline characteristics

	All randomised patients (analysis population 1)	Patients given a drug recommendation (analysis population 2)
	N=XXX	N = xxx
Demographics		
Age (years): Mean (SD)	xx (xx)	xx (xx)
Female sex at birth	xx (xx)	xx (xx)
White ethnicity	xx (xx)	xx (xx)
In a paid job	xx (xx)	xx (xx)
Marital Status	xx (xx)	xx (xx)
Married	xx (xx)	xx (xx)
Widowed	xx (xx)	xx (xx)
Divorced	xx (xx)	xx (xx)
Separated	xx (xx)	xx (xx)
Co-habiting	xx (xx)	xx (xx)
Single	xx (xx)	xx (xx)
Index of multiple deprivation (1 to 32844): Mean (SD)	xx (xx)	xx (xx)
Quintile 1: IMD 1 to 6568	xx (xx)	xx (xx)
Quintile 2: IMD 6569 to 13137	xx (xx)	xx (xx)
Quintile 3: IMD 13138 to 19706	xx (xx)	xx (xx)
Quintile 4: IMD 19707 to 26275	xx (xx)	xx (xx)
Quintile 5: IMD 26276 to 32844	xx (xx)	xx (xx)
Stratifying variables in the randomisation process		
Site		
Stoke	xx (xx)	xx (xx)
Oxford	xx (xx)	xx (xx)
Portsmouth	xx (xx)	xx (xx)
Wolverhampton	xx (xx)	xx (xx)
Trial outcome measures (where measured at baseline)		
Modified brief illness perceptions questionnaire: Mean (SD)		
Timeline (0-10)	xx (xx)	xx (xx)
Consequences (0-10)	xx (xx)	xx (xx)
Personal control (0-10)	xx (xx)	xx (xx)
Treatment control (0-10)	xx (xx)	xx (xx)
Emotional representation (0-10)	xx (xx)	xx (xx)
Understanding (0-10)	xx (xx)	xx (xx)
Illness coherence (0-10)	xx (xx)	xx (xx)
Concern (0-10)	xx (xx)	xx (xx)
Causal/identity (0-10)	xx (xx)	xx (xx)
Belief Medicines Questionnaire (BMQ) - general subscale: Mean (SD)	xx (xx)	xx (xx)
Weight (kgs): Mean (SD)	xx (xx)	xx (xx)

Body-mass index (BMI) (kg/m^2): Mean (SD)	xx (xx)	xx (xx)
Categorised BMI		
Underweight: BMI <18.5 kg/m^2	xx (xx)	xx (xx)
Normal weight: BMI>=18.5kg/m^2&<24.9 kg/m^2	xx (xx)	xx (xx)
Overweight: BMI >= 24.9 kg/m^2 & < 29.9 kg/m^2	xx (xx)	xx (xx)
Obese: BMI >= 29.9 kg/m^2	xx (xx)	xx (xx)
Current smoker	xx (xx)	xx (xx)
If a non-smoker, previous history of smoking	xx (xx)	xx (xx)
Alcohol consumption		
Daily or most days	xx (xx)	xx (xx)
Once or twice a week	xx (xx)	xx (xx)
Once or twice a month	xx (xx)	xx (xx)
Once or twice a year	xx (xx)	xx (xx)
Never	xx (xx)	xx (xx)
Baseline fractures, falls, treatment use and other health conditions		
Location of baseline facture		
Upper arm or shoulder (humerus)	xx (xx)	xx (xx)
Forearm or wrist (distal radius)	xx (xx)	xx (xx)
Hip (femur)	xx (xx)	xx (xx)
Pelvis	xx (xx)	xx (xx)
Spine (vertebral)	xx (xx)	xx (xx)
Other	xx (xx)	xx (xx)
Unsure	xx (xx)	xx (xx)
Falls in the 6-months prior to baseline	xx (xx)	xx (xx)
Biological mother or father broke their hip	xx (xx)	xx (xx)
Any previous use of osteoporosis medicine	xx (xx)	xx (xx)
Health conditions		
Type I (insulin dependent) diabetes	xx (xx)	xx (xx)
Osteogenesis imperfecta	xx (xx)	xx (xx)
Overactive thyroid	xx (xx)	xx (xx)
Premature menopause (before the age of 45)	xx (xx)	xx (xx)
Malabsorption	xx (xx)	xx (xx)
Chronic liver disease	xx (xx)	xx (xx)
Rheumatoid arthritis	xx (xx)	xx (xx)

Figures are numbers (percentages) unless otherwise stated. IQR = Inter-quartile range; SD = standard deviation

Table 13.2: Clinical effectiveness of secondary outcomes collected on patients given an osteoporosis drug recommendation (N = X)

Outcome measure	2-weeks	3-months
Primary outcome		
Decisional conflict scale: (0-100)		
Usual care: Mean (SD)	xx (xx)	N/A
iFraP: Mean (SD)	xx (xx)	N/A
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	xx (xx, xx)	N/A
Usual care vs iFraP: Adjusted$^{\beta}$ mean difference (95% CI)	xx (xx, xx)	N/A
Secondary outcomes		
Beliefs about medicines (BMQ : specific subscale): (10-50)		
Usual care: Mean (SD)	N/A	xx (xx)
iFraP: Mean (SD)	N/A	xx (xx)
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	N/A	xx (xx, xx)
Satisfaction with medicines information (SIMS): (0-17)		
Usual care: Mean (SD)	xx (xx)	N/A
iFraP: Mean (SD)	xx (xx)	N/A
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	xx (xx, xx)	N/A
Self-reported: medicine initiation or intention to initiate		
Usual care: N (%)	xx (xx)	xx (xx)
iFraP: N (%)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted$^{\alpha}$ odds ratio (95% CI)	xx (xx, xx)	xx (xx, xx)
Self-reported: adherence (MARS-5): (5-25)		
Usual care: Mean (SD)	N/A	xx (xx)
iFraP: Mean (SD)	N/A	xx (xx)
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	N/A	xx (xx, xx)
Self-reported: medicine persistence/discontinuation		
Usual care		
Did not start taking medication: N (%)	N/A	xx (xx)
Discontinued with medication: N (%)	N/A	xx (xx)
Continued medication: N (%)	N/A	xx (xx)
iFraP		

Did not start taking medication: N (%)	N/A	xx (xx)
Discontinued with medication: N (%)	N/A	xx (xx)
Continued medication: N (%)	N/A	xx (xx)
Usual care vs iFraP: Adjusted[α] odds ratio (95% CI)	N/A	xx (xx, xx)
Medical record review: medication initiation		
Usual care: N (%)	N/A	xx (xx)
iFraP: N (%)	N/A	xx (xx)
Usual care vs iFraP: Adjusted[α] odds ratio (95% CI)	N/A	xx (xx, xx)
Medical record review: medicine persistence/discontinuation		
Usual care	N/A	xx (xx)
Did not start taking medication: N (%)	N/A	xx (xx)
Discontinued with medication: N (%)	N/A	xx (xx)
Continued medication: N (%)		
iFraP	N/A	xx (xx)
Did not start taking medication: N (%)	N/A	xx (xx)
Discontinued with medication: N (%)	N/A	xx (xx)
Continued medication: N (%)		
Usual care vs iFraP: Adjusted[α] odds ratio (95% CI)	N/A	xx (xx, xx)

[α] Adjusted for FLS site, age (years), sex at birth (male, female), and the index of multiple deprivation. β Adjusted for FLS site, age (years), sex at birth (male, female), index of multiple deprivation, and the treating FLS clinician. CI = confidence interval

Table 13.3: Clinical effectiveness of secondary outcomes collected on all randomised patients (N =X)

Outcome measure	2-weeks	3-months
Patient-Professional Interaction Questionnaire (PPIQ): (16-80)		
Usual care: Mean (SD)	xx (xx)	N/A
iFraP: Mean (SD)	xx (xx)	N/A
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	xx (xx, xx)	N/A
Satisfaction with amount of verbal information: (6-30)		
Usual care: Mean (SD)	xx (xx)	N/A
iFraP: Mean (SD)	xx (xx)	N/A
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	xx (xx, xx)	N/A
Satisfaction with consultation experience		
Usual care:	xx (xx)	N/A
Strongly disagree: N (%)	xx (xx)	N/A
Disagree: N (%)	xx (xx)	N/A
Neither agree nor disagree: N (%)	xx (xx)	N/A
Agree: N (%)	xx (xx)	N/A
Strongly agree: N (%)	xx (xx)	N/A
iFraP:		
Strongly disagree: N (%)	xx (xx)	N/A
Disagree: N (%)	xx (xx)	N/A
Neither agree nor disagree: N (%)	xx (xx)	N/A
Agree: N (%)	xx (xx)	N/A
Strongly agree: N (%)	xx (xx)	N/A
Usual care vs iFraP: Adjusted$^{\alpha}$ odds ratio (95% CI)	xx (xx, xx)	N/A
Satisfaction with the amount of written information: (3-15)		
Usual care: Mean (SD)	N/A	xx (xx)
iFraP: Mean (SD)	N/A	xx (xx)
Usual care vs iFraP: Adjusted$^{\alpha}$ mean difference (95% CI)	N/A	xx (xx, xx)
Illness perceptions: Timeline (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)

Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Consequences (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Personal control (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Treatment control (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Emotional Representations (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Understanding (0-100) (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: Illness coherence (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: concern (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)
Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)
Illness perceptions: causal/identity (0-10)		
Usual care: Mean (SD)	xx (xx)	xx (xx)
iFraP: Mean (SD)	xx (xx)	xx (xx)

Usual care vs iFraP: Adjusted[α] mean difference (95% CI)	xx (xx, xx)	xx (xx, xx)

[α] Adjusted for FLS site, age (years), sex at birth (male, female), and the index of multiple deprivation. CI = confidence interval

14 Outline tables and figures (supplementary material)

Table 14.1: Baseline factors that may influence ease of communication in the consultation

	All randomised patients (analysis population 1)	Patients given a drug recommendation (analysis population 2)
	N=XXX	N = xxx
English as main language	xx (xx)	xx (xx)
Understand English: very well	xx (xx)	xx (xx)
Speak English: very well	xx (xx)	xx (xx)
Access to:		
Landline telephone	xx (xx)	xx (xx)
Desktop or laptop computer with internet access	xx (xx)	xx (xx)
Basic mobile phone (phone calls and texts only)	xx (xx)	xx (xx)
iPad or other tablet	xx (xx)	xx (xx)
Smartphone (can access the internet)	xx (xx)	xx (xx)
Frequency of internet use		
Never	xx (xx)	xx (xx)
Sporadically (less than 1 day/week)	xx (xx)	xx (xx)
Regularly (1–3 days/week)	xx (xx)	xx (xx)
Frequently (4–6 days/ week)	xx (xx)	xx (xx)
Daily	xx (xx)	xx (xx)
Help required to read instructions, pamphlets, or other written material from your doctor or pharmacy		
Never	xx (xx)	xx (xx)
Rarely	xx (xx)	xx (xx)
Sometimes	xx (xx)	xx (xx)
Often	xx (xx)	xx (xx)
Always	xx (xx)	xx (xx)
Hearing (with a hearing aid if using)		
Excellent	xx (xx)	xx (xx)
Very good	xx (xx)	xx (xx)
Good	xx (xx)	xx (xx)
Fair	xx (xx)	xx (xx)
Poor	xx (xx)	xx (xx)
Eyesight (using glasses or corrective lenses if using)		
Excellent	xx (xx)	xx (xx)
Very good	xx (xx)	xx (xx)
Good	xx (xx)	xx (xx)
Fair	xx (xx)	xx (xx)

Poor or registered blind	xx (xx)	xx (xx)

Figures are numbers (percentages)

Table 14.2: External validity of the trial sample

Characteristic	Sent a study invite N= XX	Provided consent to further contact N= XX	Randomised to the study N= XX
Age at date of study invite	xx (xx)	xx (xx)	xx (xx)
Female sex at birth	xx (xx)	xx (xx)	xx (xx)
Index of multiple deprivation	Data not collected	xx (xx)	xx (xx)

Table 14.3: Lifestyle changes at 3-month follow-up

	Usual care	iFraP intervention
Reduced alcohol consumption	xx (xx)	xx (xx)
Stopped smoking	xx (xx)	xx (xx)
Gained weight if underweight	xx (xx)	xx (xx)
Increased physical activity	xx (xx)	xx (xx)

Figures are numbers and percentages. <<Add in footnote to explain how each factor has been defined>>.

Table 14.4: Patient perception of the risk of breaking a bone at baseline and 2-week follow-up

Usual care						
	2-week follow-up					
	Much lower	A little lower	About the same	A little higher	Much Higher	Total
Baseline						
Much lower	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
A little lower	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
About the same	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
A little higher	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
Much higher	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
Total	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**
iFraP						
	2-week follow-up					
	Much lower	A little lower	About the same	A little higher	Much Higher	Total
Baseline						
Much lower	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
A little lower	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
About the same	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
A little higher	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
Much higher	x (xx)	x (xx)	x (xx)	x (xx)	x (xx)	**x (xx)**
Total	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**

Perception of risk determined from the question "*Compared to other people of the same sex and age as you, do you consider your chance of breaking a bone is….".* Orange shading indicates patients whose perception of risk has reduced at the 2-week follow-up. Blue shading indicates patients whose perception of risk has increased at the 2-week follow-up. Green shading indicates patients whose perception of risk remains unchanged after taking part in the trial.

Table 14.5: Alignment of patient perception of fracture risk and risk as determined by FRAX score

	Patient perception of fracture risk compared to other people of the same age and sex			
Usual care				
	Lower	About the same	Higher	Total
FRAX risk score				
Low	x (xx)	x (xx)	x (xx)	**x (xx)**
Middle	x (xx)	x (xx)	x (xx)	**x (xx)**
High	x (xx)	x (xx)	x (xx)	**x (xx)**
Total	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**
iFraP				
	Lower	About the same	Higher	Total
FRAX risk score				
Low	x (xx)	x (xx)	x (xx)	**x (xx)**
Middle	x (xx)	x (xx)	x (xx)	**x (xx)**
High	x (xx)	x (xx)	x (xx)	**x (xx)**
Total	**x (xx)**	**x (xx)**	**x (xx)**	**x (xx)**

Patient perception of risk determined from the question *"Compared to other people of the same sex and age as you, do you consider your chance of breaking a bone is…."*. Orange shading indicates patients whose perception of risk is less than their risk as calculated by their FRAX score. Blue shading indicates patients whose perception of risk is greater than their risk as calculated by their FRAX score. Green shading indicates patients whose perception of risk is aligned with their risk as calculated by their FRAX score. FRAX algorithm: (https://www.sheffield.ac.uk/FRAX/)

Table 14.6: Worry about falls and fractures at 2-week follow-up for patients given an osteoporosis drug recommendation

	Usual care (N= xx)	iFraP intervention N = (xx)
Worry about falls in the next 2 months		
Increased by x number of points		
X = 1	xx (xx)	xx (xx)
X = 2	xx (xx)	xx (xx)
X = 3	xx (xx)	xx (xx)
X = 4	xx (xx)	xx (xx)
X = 5	xx (xx)	xx (xx)
Decreased by x number of points		
X = 1	xx (xx)	xx (xx)
X = 2	xx (xx)	xx (xx)
X = 3	xx (xx)	xx (xx)
X = 4	xx (xx)	xx (xx)
X = 5	xx (xx)	xx (xx)
Worry about fractures in the next 2 months		
Increased by x number of points		
X = 1	xx (xx)	xx (xx)
X = 2	xx (xx)	xx (xx)
X = 3	xx (xx)	xx (xx)
X = 4	xx (xx)	xx (xx)
X = 5	xx (xx)	xx (xx)
Decreased by x number of points		
X = 1	xx (xx)	xx (xx)
X = 2	xx (xx)	xx (xx)
X = 3	xx (xx)	xx (xx)
X = 4	xx (xx)	xx (xx)
X = 5	xx (xx)	xx (xx)

Figures are numbers (percentages). Worry is measure at baseline and at the 2-week follow-up on a 6-point scale with 1 = Not worried at all; 6 = Very worried

Table 14.7: Specific osteoporosis values measured at 2-week follow-up

	Usual care (N= xx)	iFraP intervention N = (xx)
Importance of osteoporosis medicine benefit		
Strengthening bone		
Not at all	xx (xx)	xx (xx)
Slightly	xx (xx)	xx (xx)
Moderately	xx (xx)	xx (xx)
Very	xx (xx)	xx (xx)
Extremely	xx (xx)	xx (xx)
Lowering chance of future fractures & protecting spine		
Not at all	xx (xx)	xx (xx)
Slightly	xx (xx)	xx (xx)
Moderately	xx (xx)	xx (xx)
Very	xx (xx)	xx (xx)
Extremely	xx (xx)	xx (xx)
Maintaining independence		
Not at all	xx (xx)	xx (xx)
Slightly	xx (xx)	xx (xx)
Moderately	xx (xx)	xx (xx)
Very	xx (xx)	xx (xx)
Extremely	xx (xx)	xx (xx)
Factors deterring osteoporosis medicine use		
Concerns about rare long-term issues with medicines such as the jawbone problem		
Not at all	xx (xx)	xx (xx)
Slightly	xx (xx)	xx (xx)
Moderately	xx (xx)	xx (xx)
Very	xx (xx)	xx (xx)
Extremely	xx (xx)	xx (xx)
Concerns about common side-effects with medicines such as indigestion and reflux		
Not at all	xx (xx)	xx (xx)
Slightly	xx (xx)	xx (xx)
Moderately	xx (xx)	xx (xx)
Very	xx (xx)	xx (xx)
Extremely	xx (xx)	xx (xx)

Figures are numbers (percentages)

Table 14.8: Exploratory subgroup analyses for the Decisional Conflict Scale (DCS) at 2-week follow-up

	Mean (SD) DCS	Interaction (95% CI)
Mode of consultation		
Face to face: N = X		
Usual care	xx(xx)	0
iFraP	xx(xx)	
Telephone: N = X		
Usual care	xx(xx)	xx (xx, xx)$^{\beta}$
iFraP	xx(xx)	
Bone health record		
Yes: N = X		
Usual care	xx(xx)	0
iFraP	xx(xx)	
No: N = X		
Usual care	xx(xx)	xx (xx, xx)$^{\beta}$
iFraP	xx(xx)	
Health literacy$^{\alpha}$		
No diffculty: N = X		
Usual care	xx(xx)	0
iFraP	xx(xx)	
Some difficulty: N =X		
Usual care	xx(xx)	xx (xx, xx)$^{\beta}$
iFraP	xx(xx)	
Sex at birth		
Male: N = X		
Usual care	xx(xx)	0
iFraP	xx(xx)	
Female: N = X		
Usual care	xx(xx)	xx (xx, xx)$^{\beta}$
iFraP	xx(xx)	

DCS = Decisional conflict scale. SD = standard deviation, CI = confidence interval. α Defined at baseline using the question: *How often do you need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy? Never vs Rarely, Sometimes, Often, Always.* β Adjusted for FLS site, age (years), sex at birth (male, female), and the index of multiple deprivation.

Table 14.9: Side effect symptoms in patients taking osteoporosis medicines since their FLS appointment (N = XX)

	Patient reports symptom	Patient reports symptom *and believes it is caused by osteoporosis*	Patient reports symptom *and believes it is caused by osteoporosis medication*	Patient reports symptom *but is unclear on the cause/believes cause is unrelated to osteoporosis*
Usual care				
Muscle pain	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Indigestion, acid or reflux	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Headache	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Flu like symptoms	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Nausea	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Dizziness	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Heart palpitations	xx (xx)	xx (xx)	xx (xx)	xx (xx)
iFraP				
Muscle pain	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Indigestion, acid or reflux	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Headache	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Flu like symptoms	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Nausea	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Dizziness	xx (xx)	xx (xx)	xx (xx)	xx (xx)
Heart palpitations	xx (xx)	xx (xx)	xx (xx)	xx (xx)

Figures are numbers (percentages). <<*Add in a description of any other symptoms that patients report in the "other" box*>>

15 Appendices

15.1 Pre-planned adaptations to the imputation strategy

Numerical issues, failure, and breakdown of the multiple imputation algorithm can arise, particularly when there are many variables to include in the imputation model (Nguyen et al. 2021). If this does arise, we plan to use the strategy below (sequentially) to explore how the imputation model can be adapted to ensure that it can be applied to the data.

15.1.1 Perfect prediction

Perfect prediction can arise when multiple categorical variables are included in the imputation model. This would be addressed by adding the STATA "augment" option to the imputation model – a procedure that works by adding in additional "pseudo-observations" to prevent the outcome being perfectly predicted (Nguyen et al. 2021).

15.1.2 Ordinal variables

Ordinal variables can be challenging to include in an imputation model due to the number of categories they contain. If, after inspection of the imputation model, it appears that the reason why the imputation model will not run is due to the inclusion of too many ordinal variables, we will use the STATA "ascontinuous" option for the ordinal variables. This imputes the ordinal outcomes using ordinal regression, but, when these outcomes are included as predictor variables in the imputation model for other outcomes, they are assumed to be continuous variables, rather than categorical, to reduce the number of degrees of freedom in the imputation model (StataCorp. 2022). We will only use this approach for ordinal outcomes that are measured using a relatively large number of the response categories (>= 4).

15.1.3 Number of nearest neighbours (k) in the predictive mean matching (PMM) models

Kleinke 2018, highlight that there is a trade-off when considering the number of nearest neighbours (k) to include in the PMM model: if k is too small a single participant's data could be repeatedly chosen as a donor in the imputation model, which would underestimate model standard errors, whereas if k is too large might results in inadequate donors and implausible imputations, hence biased inferences.

We have used the recommendation by Morris et al. 2014 to set the value of k in the imputation model to be 10. If this decision means that the imputation model breaks down when we fit it to our data, we will re-run the imputation model, firstly with k=5 and then secondly with k = 15 to see if these changes enable the imputation model to run in our data. We will try K=5, before K=15, as the former is preferred default value for K used in the alternative statistical software packages of SAS and R (Kleinke 2018).

15.1.4 Merging of categories for outcomes with categorical response options

For categorical outcomes, it may be, that prior to analysis, merging of categories is required due to a small number of patients in certain response options. If this is the case, the merged version of the variable would be included in the imputation model (rather than the version with all response options) to simplify the model and encourage model convergence

15.1.5 Dropping variables from the imputation model

It is recognised that we will have many variables to include as predictors in the imputation model. If, after the strategies described above have been employed, the imputation model still does not converge we will consider dropping the Illness perceptions subscales from the model. These measures are collected at more than one time-point, so it is possible to analyse them in a mixed-models framework where the assumption that the data are missing at random is handled directly in the model rather than requiring multiple imputation

15.1.6 Adapting the imputation model

We have chosen to use MICE as our imputation method, however, other imputation models exist, such as Multi-Variate Normal Imputation (MVNI), which could be used as an alternative approach (Nguyen et al. 2021). Therefore, if our MICE imputation model is unsuccessful, we will explore changing the imputation method to MVNI, to see if we can successfully impute the data using this method.