Pain-LESS

Characterisation of **Pain** in patients with musculoskeletal disease: a prospective, **L**ongitudinal, observational study with an **E**mbedded feasibility window of opportunity **S**leep **S**tudy

Statistical Analysis Plan

Version 1.1 – 29th April 2025

Previous versions:
Version 1 – 17th March 2025

Based on Protocol Version 5 – 7th February 2023

Trial registration: NCT05962138

Note: *We thank OCTRU for their support in the formulation and structure of the SAP*

CONTENTS

1. INTRODUCTION

This document details the proposed data presentation and analysis for the main paper(s) and final study reports from the *NIHR*-funded single-centre Randomised Controlled Trial of digital cognitive behavioural therapy for insomnia (dCBT-I; *Sleepio*) in the treatment of fibromyalgia (PainLESS). The results reported in these papers should follow the strategy set out here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial. This document follows published guidelines regarding the content of statistical analysis plans for clinical trial(Gamble et al. 2017).

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

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

1.1 Key personnel

List of key people involved in the drafting and reviewing this SAP, together with their role in the trial and their contact details.

Author(s):

- **Dr Eoin Kelleher** *(Investigator)*

Reviewers:

- **Dr Daphne Kounali** *(Senior statistician)*
- **Dr Anushka Irani** *(Investigator)*
- **Amanda Wall** *(Investigator)*
- **Prof Ben Seymour** *(Chief Investigator)*

Approvers:

- **Prof Ben Seymour** *(Chief Investigator)*

1.2 Changes from previous version of SAP

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

Version number Issue date	Author of this issue	Protocol Version & Issue date	Significant changes from previous version together with reasons
V1.0_17Mar2025	Eoin Kelleher	Protocol_V5.0_23Feb2023	Not applicable as this is the 1st issue
V1.1_29Apr2025	Eoin Kelleher	Protocol_V5.0_23Feb2023	Language changes to enhance clarity

2. BACKGROUND AND OBJECTIVES

Fibromyalgia is a debilitating and poorly understood condition characterised by chronic widespread pain, fatigue, cognitive impairment, and sleep disturbance. Affecting approximately 5% of the UK population, fibromyalgia imposes a significant socioeconomic burden, with many patients experiencing reduced productivity and increased reliance on healthcare services(Sarzi-Puttini et al. 2020; Fayaz et al. 2016; Soni et al. 2020). Within five years of diagnosis, 25% of fibromyalgia patients are unable to work, underscoring the need for effective interventions(Spaeth 2009; Soni et al. 2020).

Sleep disturbance is a hallmark feature of fibromyalgia, reported by up to 90% of patients, and it significantly impacts quality of life, pain management, and cognitive functioning ("fibrofog")(Clauw 2014; Choy 2015). This vicious cycle between disrupted sleep, pain, and cognitive dysfunction diminishes quality of life and functional capacity(Clauw 2014; Choy 2015). Despite its near-universal presence in fibromyalgia, sleep dysfunction remains relatively under investigated as a therapeutic target, and accessible, scalable treatment options are lacking.

Patient priorities, as outlined by the James Lind Alliance and recent NICE guidelines (NG193('National Institute for Health and Care Excellence: Guidelines' 2021)), have emphasised the need to develop and evaluate sleep interventions in chronic pain populations(Fitzcharles et al. 2017).

Cognitive Behavioural Therapy for Insomnia (CBT-I) is the gold-standard non-pharmacological treatment for chronic insomnia, with robust evidence supporting its efficacy in improving sleep quality and quality of life(Riemann et al. 2023). Beyond its benefits for sleep, CBT-I has shown potential for addressing pain, fatigue, and cognitive symptoms in chronic pain conditions(Tang et al. 2015; Salazar-Méndez et al. 2024; Kyle et al. 2020). However, conventional face-to-face CBT-I often faces logistical and financial barriers, limiting its accessibility for fibromyalgia patients. Digital CBT-I (dCBT-I), delivered via online platforms, offers a scalable, cost-effective alternative, with evidence demonstrating comparable efficacy to in-person therapy(Zachariae et al. 2016).

Despite its promise, the application of dCBT-I in fibromyalgia is underexplored. Several pilot studies in other chronic pain conditions, such as osteoarthritis and migraine, suggest potential benefits, but large-scale trials in fibromyalgia are lacking(Crawford et al. 2020; Whibley et al. 2022). Moreover, the mechanisms by which dCBT-I may improve sleep and quality of life in fibromyalgia, and alleviate symptoms such as cognitive dysfunction, remain unclear. This research gap highlights an opportunity to evaluate the feasibility and impact of dCBT-I in fibromyalgia and to explore its underlying mechanisms.

In this trial, we aim to address these gaps by conducting a randomised controlled trial of dCBT-I, delivered via *Sleepio*, in patients with fibromyalgia. The study examines a comprehensive range of outcomes, including quality of life, sleep quality, cognitive performance, and neuroimaging measures. By embedding the trial within a larger observational cohort, the study aligns with real-world clinical settings and ensures efficient data utilisation(Kim, Flory, and Relton 2018).

The aim of this pragmatic randomised controlled superiority trial is to evaluate the clinical and cost-effectiveness of dCBT-I versus sleep hygiene advice for improving quality of life in adults with fibromyalgia. The estimand for the primary objective (including the analysis of the primary outcome) is described in **Table 1**.

Table 1. *Estimand-to-analysis table*

Primary objective: To quantify and draw inferences on observed differences in fibromyalgia-related quality of life for adults with fibromyalgia between dCBT-I versus sleep hygiene advice at 3-months post-randomisation

Estimand: The difference in fibromyalgia-related quality-of-life scores between those treated with dCBT-I and sleep hygiene, and those given sleep hygiene advice only, 3-months following randomisation, irrespective of sleep therapy technique used and any unforeseen difficulties making the delivery of either randomised intervention impossible or difficult and/or receiving any other interventions that are part of the standard of care.	
Treatment: dCBT-I and sleep hygiene advice vs. sleep hygiene advice only	
ESTIMAND	***ANALYSIS***
Target population: Adults aged over 18 with fibromyalgia according to 2016 revised American College of Rheumatology criteria(Wolfe et al. 2016) who are patients of Oxford University Hospitals NHS Foundation Trust or Connect Health community musculoskeletal health, and who have concomitant sleep disturbance (measured used 2-item Sleep Conditions Indicator, SCI-2)	***Analysis set:*** All randomised participants. Participants assigned to dCBT-I and sleep hygiene advice will be the active treatment group. Participants assigned to the sleep hygiene only group will be the comparator group.
Variable: Fibromyalgia-related quality of life measured on the Fibromyalgia Impact Questionnaire Revised (FIQR) at 3-months post-randomisation	***Outcome measure:*** Continuous outcome scale where higher scores indicate worse quality of life, at 3-months post-randomisation
Handling of intercurrent events **Premature treatment discontinuation/non-engagement:** All follow-up will be included regardless of attribution to treatment under a treatment policy strategy. **Switching Treatment arm:** Participants who receive a different intervention to that which they were allocated will be analysed according to their randomised allocation under a principal stratum strategy	***Handling of missing data*** Multilevel modelling of outcome scores over time imputes implicitly intermittent missing data over the post-randomisation period of follow-up. The underlying assumption is missing at random (MAR) conditional on all other outcome scores and minimisation factors. This is the main approach to handling missing data. **Sensitivity analysis:** The MAR assumption can be strengthened by conditioning on additional baseline covariates predictive of the outcome and missingness. Missing data arising from participants discontinuing in the study prior to the end of follow-up under a NMAR (Not Missing at random) informative missing data mechanism will be explored using the extended Hausman-Wise-Diggle-Kenward selection model when drop out depends on outcome scores prior to drop out and where this is jointly modelled with the longitudinal outcome(Diggle and Kenward 1994).
Population-level summary measure ***Average Treatment effect (ATE):*** Mean difference in fibromyalgia-related quality of life scores at 3	***Analysis approach*** **Primary analysis: Constrained Longitudinal Data Analysis (cLDA)** – The primary analysis will use a Constrained Longitudinal Data Analysis (cLDA) model to estimate the treatment effect of

months, between the two randomised arms, as randomised, measured on the FIQR	dCBT-I on FIQR scores over time. This model ensures that baseline FIQR scores are constrained to be identical across treatment groups, reducing the need for explicit baseline adjustment in the fixed effects. The model includes fixed effects for treatment, time (categorical), and their interaction, with random intercepts at the participant level to account for repeated measures. Treatment effects will be estimated at 3, 6, and 12 months, with 95% confidence intervals. **Sensitivity Analysis: Complier Average Causal Effect (CACE):** The CACE estimand estimates the treatment effect among participants who adhered to their assigned intervention, defined as completing at least 3 of 6 dCBT-I sessions. A two-stage least squares (2SLS) instrumental variable approach will be used, leveraging randomisation as an instrument for compliance to estimate the causal effect of dCBT-I on FIQR scores among compliers. A continuous measure of compliance (0–6 sessions) will also be examined to assess a potential dose-response relationship.

3. STUDY METHODS

3.1 Trial Design/framework

PainLESS is a single-centre, prospective, randomised, superiority trial using a two-arm parallel group design in a single-centre in Oxford, UK. Adults aged 18 years or older with fibromyalgia and concomitant sleep disturbance, will be randomised in a 1:1 ratio to either receive dCBT-I and sleep hygiene advice, or sleep hygiene advice only. The nature of the treatment means that the participants cannot be blinded to the treatment received; however, the outcome assessors will be blinded to the participant's treatment allocation. This trial was nested within a larger observational cohort. Within the trial, a subgroup of eligible participants underwent a brain MRI at baseline and at 3-months post-randomisation.
The trial includes an internal pilot (phase 1) which was designed to confirm the expected rate of recruitment and feasibility of the study design, and to estimate variability of the outcome measure to inform the sample size for Phase 2. This pilot was planned to include 30 participants. The study investigators will make a decision regarding trial continuation in the event that the recruitment target for the internal pilot is not met. Otherwise, the trial would continue into the main phase (phase 2), and patients from the internal pilot would be included in the final analysis. The main trial phase was planned to be recruiting from Oxford University Hospitals NHS Foundation Trust (OUH) and Connect Health, stratified by sex to account for known imbalances in fibromyalgia prevalence, and by participation in the brain MRI sub-study to ensure a balanced sample for this sub-study. Follow-up was planned to be made in person at 3 months, and electronically (sent by e-mail or text message) for patient reported outcomes at 3 months (primary outcome time-point), 6 months and 12 months.
All adults aged 18 years or older with a diagnosis of fibromyalgia, who are patients of OUH or Connect Health, who have concomitant self-reported sleep disturbance (SCI-2) and brain fog, and who had access to a device with an internet connection humerus are potentially eligible to take part in the trial. Participants with a major neuropsychiatric condition (not including depression or anxiety), neurodegenerative disorder, epilepsy, recent surgery, regular nightshift work, regular use of sedative medications, an untreated sleep disorder apart from insomnia (e.g. obstructive sleep apnoea), who are pregnant or breastfeeding, or who are currently using psychological therapies for insomnia are not eligible. After consent has been gained, an online survey will collect baseline demographic data, fibromyalgia-related quality of life (FIQR), sleep quality (ISI), brain fog (BC-CCI), fear of movement (TSK), pain (SF-36), and health-related quality of life using the EuroQoL EQ-5DY. At a site visit, a trial investigator will collect information on sleep (actigraphy and sleep diary), cognitive performance, physical activity, pain (QST), and brain MRI.

3.2 Randomisation and Blinding

Patients will be randomised after consent and collection of baseline data. Participants will be randomised on a 1:1 basis using *Sealed Envelope,* a secure electronic randomisation platform. To ensure balanced group allocation, randomisation will be stratified by sex and completion of a brain MRI in the neuroimaging sub study, with participants assigned in blocks of 10. Outcome assessors will be blinded to treatment allocation for all study visits and communications with participants, and investigators carrying out analysis will also be blinded. Due to the nature of the intervention, participants were unable to be blinded.

To maintain allocation concealment, all study investigators will not have access to future group allocations and are unable to influence randomisation. The treatment allocation will be communicated to participants by an independent research nurse who is not involved in study visits or outcome assessments via email. This email will consist of scripted phrases and contained instructions for *Sleepio* access. In addition, the research nurse will act as an intermediary for any queries regarding the treatment or group allocation, and will relay questions regarding the intervention to the investigators anonymously to preserve blinding.

Due to the nature of the intervention, blinding participants was not possible. However, the outcome assessors will remain blinded to treatment allocation throughout the trial. Participants were instructed not to disclose their group assignment to study investigators during follow-up visits to maintain blinding. Participants cannot be blinded to their treatment. The outcome data will be collected directly from the participant. Outcome assessors will be blinded to the participant's treatment allocation.

Full details of the randomisation are available in se_list_242996464514928final, stored in the confidential statistical section of the TMF. No emergency randomisation plan was developed for this study.

3.3 Sample Size

In the pilot phase (phase 1), the mean FIQR at baseline was 57.69 (SD=18.69). At the 3-month post-treatment follow-up, for the overall group, the mean FIQR was 52.24 (SD=19.06). The correlation between baseline and 3-month follow-up FIQR scores was high (r=0.86; 95%CI 0.73 to 0.93).

Using these values, the sample size was estimated to detect a minimum clinically-important difference (MCID) of 14% improvement in FIQR at 3 months(Bennett, Bushmakin, et al. 2009), corresponding to an improvement of 8.08 (Cohen's d=0.43) in this study, equating to a moderate effect size. Assuming a correlation of r=0.75 (the lower bound of the 95% CI for correlation), significance level of 0.05 and 80% power, the estimated sample size required to detect a 14% improvement in FIQR amongst participants in the Sleepio arm compared to the standard care arm using an ANCOVA is 78 participants (39 per group). This calculation includes an adjustment for a 3.33% loss-to-follow-up rate.

The R code used to estimate the sample size is provided below:

```
# Inputs
alpha <- 0.05              # Significance level
power <- 0.80              # Power
percent_improvement <- 0.14  # 14% MCID based on literature
correlation <- 0.75         # Correlation (r) between baseline and outcome based on pilot sample
loss_to_follow_up <- 0.033  # Loss-to-follow-up based on pilot sample
fiqr_baseline_mean <- 57.69444  # Baseline mean based on pilot sample
fiqr_baseline_sd <- 18.68821    # Baseline standard deviation based on pilot sample

# Calculate the mean difference for 14% improvement (MCID)
mean_difference <- fiqr_baseline_mean * percent_improvement

# Calculate Cohen's d for MCID
cohen_d <- mean_difference / fiqr_baseline_sd

# Function to calculate sample size for t-test
calculate_sample_size_ttest <- function(effect_size, alpha, power) {
  z_alpha <- qnorm(1 - alpha / 2)  # Z-value for two-tailed test
  z_beta <- qnorm(power)          # Z-value for power
  n_per_group <- ((z_alpha + z_beta)^2 * 2) / (effect_size^2)
  return(ceiling(n_per_group))
}

# Calculate sample size for t-test
n_ttest <- calculate_sample_size_ttest(cohen_d, alpha, power)

# Adjust sample size using N_adjusted = N * (1 - r^2)
n_adjusted <- n_ttest * (1 - correlation^2)

# Combine groups (total sample size)
n_total_adjusted <- ceiling(2 * n_adjusted)

# Account for 10% loss to follow-up
n_adjusted_with_loss <- ceiling(n_adjusted / (1 - loss_to_follow_up))

# Total adjusted sample size (with loss-to-follow-up)
n_total_adjusted <- 2 * n_adjusted_with_loss
```

3.4 Statistical Interim Analysis, Data Review and Stopping guidelines

No Data Safety and Monitoring Committee (DSMC) was set up for this trial due to the low-risk nature of the intervention studied. There are no formal interim analyses and therefore no formal stopping rules for this study.

3.5 Timing of Analysis

The final analysis of all primary and secondary endpoints will be conducted together when all recruited patients have reached their planned follow-up up to 6 months and the relevant data received, cleaned and finalised.
The trial also includes long-term follow-up from randomisation up to 12-months; the analysis of this data will be reported separately following a separate statistical analysis plan.

3.6 Blinded analysis

A blinded analysis of the data (not separated by treatment arm) will be undertaken after the pilot sample (phase 1) have been recruited to estimate the variability of the outcome measure (FIQR). In addition, a blinded analysis of the data will be undertaken prior to the final data lock to undertake data cleaning, to look into the distribution of variables, missing data distributions, and to review exclusions associated with the per protocol population.

3.7 Statistical Analysis Outline

It is anticipated that all the statistical analysis will be undertaken using R although other well-validated statistical packages, such as Stata, may be considered for statistical quality control validation analyses. All analyses will be carried out under a treatment policy strategy (that is, all participants will be analysed in the group they were randomised to regardless of actual treatment received).

The analyses will be repeated for the *per protocol* population (definition in section 4.2) as mandated by the protocol, bearing in mind that this may introduce bias by losing the benefits of randomisation and introducing ambiguity in the primary estimand of interest.

Although we have allowed for up to 5% missing data in the sample size at 3 months, we hope to minimise this by utilising data collection techniques appropriate to the fibromyalgia patient population. Before carrying out the within trial analysis, we will check the trial data for any missing data. Where possible the reasons for missing data will be ascertained and reported. The nature and pattern of the 'missingness' will be carefully considered — including whether data can be treated as missing at random (MAR). The handling missing data is detailed in section 6.3. Standard descriptive statistics will be used to describe the demographics between the treatment groups reporting means and standard deviations or medians and interquartile ranges as appropriate for continuous variables, and numbers and percentages for binary and categorical variables. All comparative outcomes will be presented as summary statistics and reported together with 95% confidence intervals and all tests will be carried out at a 5% two-sided significance level.

The Fibromyalgia Impact Questionnaire Revised (FIQR) at 3 months is the primary outcome of the study and the primary analysis will compare this between the treatment groups in a linear mixed effects method including all participant data, adjusting for the stratification factors.

Subgroups analyses will be undertaken with the aim of exploring the consistency of treatment effects across important baseline characteristics(Yusuf et al. 1991). These include:

Insomnia with objective short sleep duration: There is evidence that patients with insomnia with *objectively* short sleep duration (<6 hours) may respond differently to CBT-I(Vgontzas et al. 2013; Ballesio et al. 2019). For this subgroup analysis, an interaction between treatment allocation and baseline TST will be fitted to evaluate effect modification by sleep duration prior to randomisation. The continuous measure of TST will be used to maximise power.

Forest plots with 95% confidence interval will be used to present the subgroup analysis results.

Secondary clinical outcomes will be similarly analysed using mixed effects regression, using logistic regression for binary data and linear regression for continuous data. Any subsequent changes in this plan are strongly discouraged. If any changes to this SAP are implemented these will be documented in an updated version of this SAP. Documentation of any such changes will include the rationale for the changes, and when and by whom these were agreed.

4. STATISTICAL PRINCIPLES

4.1 Statistical Significance and Multiple Testing

There is a single pre-specified primary outcome, so there is no correction for multiple testing. A significance level of 0.05 will be used, with 95% confidence intervals reported. All secondary analyses will be considered as supporting the primary analysis and will also be analysed using a significance level of 0.05 with 95% confidence intervals. No adjustment for multiple testing will be performed. In the setting of clinical trials, multiplicity adjustment is often inappropriate as it assumes all tested hypotheses are independent and equally relevant, which is rarely the case in this setting (for example, sleep quality, sleep efficiency, subjective dyscognition, and objective sustained attention are likely to be interrelated). Instead, adjustments like the Bonferroni correction can overcorrect, reducing statistical power and potentially obscuring meaningful clinical findings, especially when endpoints are interrelated and when clinical interpretation is more relevant than strict statistical significance(Schulz and Grimes 2005).

4.2 Definition of Analysis Populations

Populations for analysis are defined as follows:

- **Intention-to-treat (ITT):** all participants analysed in their randomised groups, regardless of actual treatment received. This reflects the primary estimand of interest (treatment policy).
- **Per-Protocol (PP):** participants who received the intervention as intended will be analysed according to the treatment they actually received. Participants will be excluded from the per-protocol population if:
    - They did not receive the treatment allocated through randomisation
    - They did not fully satisfy the eligibility criteria for the study

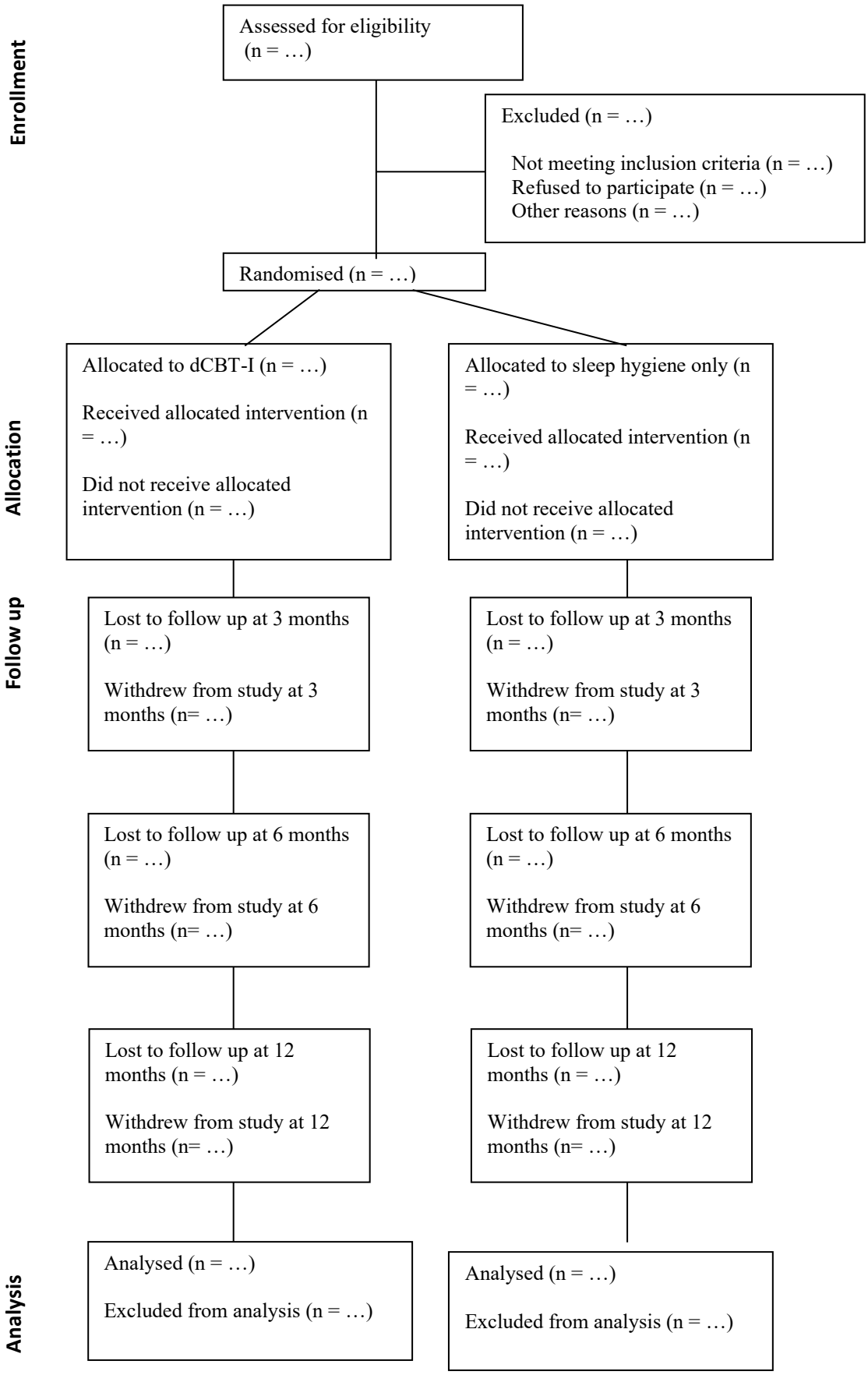
Blinded review of the protocol deviation data (not separated by treatment arm) which may affect the PP analyses will be undertaken prior to the final data lock. A summary of the characteristics of exclusions to define the per-protocol population will be reported. The motivation for any such exclusion is to examine the robustness of inferences to factors that can threaten the study's internal validity. Deviations associated with intercurrent events which occurred in the trial setting but would not be observed in practice are considered the main threat, consistent with the primary ITT estimand. If a significant number of such deviations are present e.g. involving more than >5% of participants, the results of the primary analyses may introduce ambiguity in the interpretation. In this case sensitivity analyses is undertaken and detailed in section 6.4.

5. TRIAL POPULATION AND DESCRIPTIVE ANALYSES

5.1 Representativeness of Study Sample and Patient Throughput

A CONSORT flow diagram (**Figure 1**) will be used to summarise the flow of participants through each stage of the trial, including a breakdown of the number of participants in each stage of the trial from screening through recruitment to the end of the trial. The number participants who are excluded, declined consent, withdrew or were lost to follow-up are also summarised in the flow chart.

Figure 1 The PainLESS CONSORT flow diagram



Additional expansion of reasons and timing of withdrawals from the study will be included as needed as in **Table 2** below.

Table 2: Reasons for exclusion

Reason for exclusion	**N**
Not meeting inclusion criteria	
No sleep problems	
No brain fog	
Untreated sleep disorder apart from insomnia	
Nightshift worker	
Prescribed regular sedative medications	
No fibromyalgia diagnosis	
Other medical contraindication (e.g. epilepsy)	
Other	
Total ineligible	
Eligible but refused	
Patient does not want to be part of the research project	
Participant does not want to complete the questionnaires	
Participant does not want to attend baseline study visit	
Patient decision – no reason given	
Other	
Other	
Technical difficulties	
Research staff not available/informed	
Other	
Total eligible but not randomised	
Total patients screened but not randomised	

5.2 Withdrawal from treatment and/or follow-up

The numbers and percentages of participants who are lost to follow-up or withdraw will be reported by treatment allocation for each time point until the end of follow-up at 12 months post-randomisation. Reasons for withdrawal will also be summarised by treatment allocation in **Table 3a** and **Table3b**.

Table 3a: Summary of withdrawals and losses to follow from baseline to 12 months' post-randomisation. For the purpose of this report, a participant is considered to be lost to follow-up at a particular visit if the primary outcome assessments following that visit are all missing.

	dCBT-I and sleep hygiene				**Sleep hygiene only**			
Totals	**Withdrawals**	**%**	**Lost-to-follow-up**	**%**	**Withdrawals**	**%**	**Lost-to-follow-up**	**%**
Baseline								
3 months								
6 months								
12 months								

Table 3b: Summary of reasons for withdrawal by treatment allocation

Reason for withdrawal	**dCBT-I and sleep hygiene**		**Sleep hygiene only**		**Total**	
	N	**%**	**N**	**%**	**N**	**%**
Total withdrawn						
Patient does not want to be part of study						
Participant does not want to complete questionnaires						
Participant does not want to attend study visit						
No reason given						
Other reason						

5.3 Baseline Characteristics

Baseline characteristics will be reported by treatment group, including the stratification/minimisation factors and important prognostic, demographic and clinical covariates, as outlined below in **Table 4**.

Numbers (with percentages) for binary and categorical variables and mean (and standard deviation), or median (with lower and upper quartiles) for continuous variables will be presented; there will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any baseline variable.

Table 4. Baseline characteristics of participants

	dCBT-I		Sleep Hygiene		Total	
Continuous	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age, years						
FIQR						
SF36 Bodily Pain Scale						
ISI						
BC-CCI						
Sleep duration, hours [actigraphy]						
Time in bed, hours [actigraphy]						
Sleep efficiency, % [actigraphy]						
PHQ-9						
GAD-7						
TSK						
EQ-5D-5L utility						
EQ-5D-5L VAS						
Categorical	**N**	**%**	**N**	**%**	**N**	**%**
Gender						
Male						
Female						
Education						
Higher Degree						

No higher degree						
Employment						
Employed						
Not employed						
MRI scan						
Yes						
No						

5.4 Unblinding

All cases of unblinding will be listed, together with who was unblinded, reasons for unblinding and summarised (numbers, percentages).

5.5 Treatment Compliance with Details of Intervention

The intervention groups in this trial are dCBT-I and sleep hygiene advice, or sleep hygiene advice only, in adults with fibromyalgia. All participants will be given sleep hygiene advice at the baseline visit. Compliance for dCBT-I is defined by the proportion of participants who complete each of the 6 sessions of dCBT-I. A complier will be defined as a participant who completes at least 3 of the 6 sessions, as session #3 contains sleep restriction therapy, the core component of CBT-I. The numbers and percentages of participants receiving each treatment are summarised, as well as those who did not receive the allocated treatment. Reasons are given in the cases where the participants did not receive the allocated treatment (**Table 5a**). The numbers and percentages of participants in the dCBT-I arm who access, and complete, each session are summarised (**Table 5b).**

Table 5a. Compliance with allocated intervention

Treatment received	**dCBT-I and sleep hygiene**		**Sleep hygiene only**	
	N	**%**	**N**	**%**
Allocated treatment				
Other				
Reason for not receiving allocated treatment				
Participant decision				
Administrative decision				
Other				

Table 5b. Compliance with dCBT-I among intervention group

Session	**Accessed session**	**Completed session**

	N	%	N	%
Session 1				
Session 2				
Session 3				
Session 4				
Session 5				
Session 6				

5.6 Reliability

To ensure consistency, validation checks of the data will be conducted. This will include checking for duplicate records, checking the range of variable values and validating potential outliers where possible. As the data is collected electronically via *REDCap*, many of these checks will be implemented automatically as part of the data entry procedure. Calculations and processes performed by a computer program, including the construction of derived data, will be checked by hand calculations. These checks will also confirm whether the data has been imported into the statistical software correctly and will check any merging of different datasets. Clarification will be sought from the study investigators in the case of discrepancies.

For variables derived from actigraphy (e.g. total sleep time, sleep efficiency), inter-rater reliability on derived variables will be assessed on a sample from 10 overlapping participants each rated independently by two trial investigators.

For each variable, missing value codes will be checked for consistency and proportion of missing values per variable will be presented. Patterns of missing data will be explored. Where missing data imputation is used, imputed values will also be verified using the validation techniques described above.

6. ANALYSIS

The statistical methods to be used to compare groups for primary and secondary outcomes and methods for point and interval estimation are summarised below. This includes methods for additional analyses, such as adjusted analyses and subgroup analyses.

The primary estimand of interest will be a treatment policy estimand (ITT population), using data from all randomised participants as randomised. This is chosen as the primary estimand to reflect the pragmatic nature of the study, and to identify the treatment effect regardless of any intercurrent events occurring. The elements of the primary estimand are summarised above in **Table 1**.

The protocol also mandates a per-protocol analysis. The per-protocol analyses (section 4.2) will include all randomised participants who received their allocated treatment but excludes those randomised participants who were found ineligible post-randomisation or did not receive their allocated treatment.

The primary analysis and sensitivity analysis described in sections 6.2-6.4 focus on the primary estimand detailed in **Table 1** which targets the average difference in FIQR scores (ATE) between the two interventions in the combined groups randomised to receive each intervention. The resulting treatment effect represents

the average difference in FIQR scores we can expect if all participants in the target population were treated with dCBT-I instead of being treated with sleep hygiene alone.

A secondary analysis adjusting for compliance with the intervention (complier average causal effect, CACE) will also be performed using instrumental variables approach (Peugh et al. 2017).

6.1 Outcome Definitions

Primary outcome: Fibromyalgia Impact Questionnaire Revised (FIQR): measured pre-randomisation (baseline), 3 months (post-treatment follow-up), and then at 6 months and 12 months. Collected electronically via REDCap survey.

The FIQR is a validated tool based on the revised symptom impact questionnaire (SIQR), widely used in rheumatology, and assesses the impact of fibromyalgia on an individual's daily life(Bennett, Friend, et al. 2009). It includes questions on the following domains: physical functioning, pain intensity and characteristics, fatigue, sleep quality, emotional well-being (mood and anxiety), and impact on work and social interactions. It is measured on a scale from 0 to 100, where higher scores indicate worse quality of life.

Secondary outcomes:

Insomnia severity: Insomnia severity will be measured pre-randomisation (baseline), and at 3-, 6-, and 12-months via the REDCap electronic survey tool. The Insomnia Severity Index (ISI) assesses severity of insomnia symptoms and their impact, and is a widely used measure of insomnia severity in both clinical and research settings(Bastien, Vallières, and Morin 2001; Chen, Jan, and Yang 2017). It consists of seven items evaluating different aspects of insomnia, including difficulties with falling asleep, staying asleep, and waking up early; satisfaction with current sleep patterns; interference with daily functioning; quality of life; worry about sleep problems. It may be a more valid measure of sleep quality changes with CBTi compared to the PSQI (see below)(Chen, Jan, and Yang 2017). It is measured on a scale of 0 to 28, where higher scores indicate worse insomnia symptom severity.

Pittsburgh Sleep Quality Index (PSQI): The PSQI will be measured pre-randomisation (baseline), and at 3-, 6-, and 12-months via the REDCap electronic survey tool. The PSQI assesses sleep quality and patterns. It complements the ISI by providing an assessment of sleep quality across seven components: sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, daytime dysfunction, and use of sleep medications(Buysse et al. 1989).

Observed sleep metrics (actigraphy): Participants will be given an actigraphy device (CamNtech MotionWatch 8) to wear on their wrist for 7 days prior to the baseline (pre-randomisation) and 3-month visits. Alongside this, they will be asked to complete the Consensus Sleep Diary each day(Carney et al. 2012), where they documented sleep times and self-reported sleep quality. The sleep diaries are used to inform the derivation of the observed sleep metrics from actigraphy. Measures derived from actigraphy include total sleep time, time in bed, sleep efficiency, sleep onset latency, and sleep fragmentation.

Pain severity: Pain severity will be measured pre-randomisation (baseline), and at 3-, 6-, and 12-months via the REDCap electronic survey tool. The 36-item short form (SF-36) health survey assesses overall health and wellbeing across eight domains(Framework 1992): physical functioning, role limitations due to physical health, role limitations due to emotional wellbeing, fatigue, emotional wellbeing, social functioning, pain, and general health. In this trial, the Bodily Pain subscale (BPS) is used as a measure of the impact of pain on physical health. It is a composite of two items measuring pain intensity and pain interference, a measure of how much pain interferes with daily activities. Scores for each subscale are transformed to a scale of 0-100, with lower scores representing worse health status.

Subjective cognitive complaints: Cognitive symptoms will be measured pre-randomisation (baseline), and at 3-, 6-, and 12-months via the REDCap electronic survey tool. The British Columbia Cognitive Complaints Inventory (BC-CCI) assesses subjective cognitive complaints through six items assessing self-reported problems over the past seven days with concentration, memory, expressing thoughts, word finding, thinking speed, and problem-solving(Iverson and Lam 2013). Although more typically used in evaluation of patients with mood disorders, this outcome measure was selected for this trial to facilitate comparison with the SPIN trial, which evaluated the effect of dCBT-I on cognitive function in patients with insomnia(Kyle et al. 2020). It is measured on a scale of 0 to 18, where higher scores indicate worse subjective cognitive symptoms.

Cognitive testing: Cognitive testing will be performed at the baseline (pre-randomisation) and post-treatment (3-month) study visits. As one of the principal symptoms of 'fibrofog' is concentration difficulties, a test of sustained visual attention was carried out during the visit(Zhao et al. 2022). This test was selected as previous studies have found deficits in sustained attention in patients with brain-fog symptoms post-COVID, which phenotypically may resemble the cognitive symptoms seen in fibromyalgia(Zhao et al. 2022). This was performed during the visit on an Apple iPad with attached keyboard on a Google Chrome browser. The task involved pressing the spacebar when '0' appeared amidst other digits (1-9), masked by a semi-transparent grey checkerboard. Each block lasted 60 seconds, and after a practice block, 9 blocks were performed in total. The hit rate was calculated from correct hits, while false positive rate was calculated from incorrect hits. A composite measure of task accuracy, D', is derived. Reaction time (RT) for correct hits is measured as the time between the digit appearing on the screen and the spacebar being pressed. RT variability is also assessed.

Health-related quality of life: Quality of life will be measured pre-randomisation (baseline), and at 3-, and 6-months via the REDCap electronic survey tool. For the planned economic evaluation, the EuroQol-5D-5L (EQ-5D-5L) will be used to estimate quality adjusted life years (QALYs), which are measures on a scale of -0.59 to 1.0, where 1.0 is a full year of perfect health. The adult value set for the UK will be used to estimate QALYs.

Healthcare resource utilisation: healthcare utilisation will be measured pre-randomisation (baseline), and at 3-, and 6-months via the REDCap electronic survey tool. For the planned economic evaluation, participants will report their healthcare service use over the previous 3 months, including hospital visits, medical imaging, and consultations with healthcare professionals (e.g., GPs, nurses, physiotherapists). Information was also collected on medications, unpaid caregiving, social services, and the financial impact of the condition on relatives and friends, including out-of-pocket expenses for care, transportation, and medical treatments.

Magnetic Resonance Imaging (MRI): A neuroimaging sub-study will be nested within the trial, with eligible participants invited to also undergo a brain MRI at the baseline (pre-randomisation) and 3-month study visits. Neuroimaging data will be acquired on a 3T Siemens Prisma MRI scanner at the Oxford Centre for Functional MRI of the Brain (FMRIB) at the John Radcliffe Hospital in Oxford. Participants will complete whole brain structural, resting state BOLD imaging, ASL imaging, and H1-MRS of the posterior insula.

Complications: Adverse events will be measured pre-randomisation (baseline), and at 3-, and 6-months via the REDCap electronic survey tool. Participants will be asked to report any potential adverse events related to dCBT-I, including physical and psychological symptoms such as fatigue, low mood, pain, and difficulty concentrating. The will rated the severity of these symptoms and how much they interfere with normal functioning, on a scale from "Did not experience" to "Very much." Participants will also report incidents of sleepiness over the preceding 3 months, including hospital admissions, accidents, falls, or near-miss incidents, and have the opportunity to mention any other relevant events.

6.2 Analysis Methods

Analysis of Primary Estimand (ATE): A Treatment Policy Analysis

Unadjusted summary statistics of the FIQR scores will be displayed by treatment allocation using means and standard deviations at each assessment point. The primary analysis will be conducted using **Constrained Longitudinal Data Analysis (cLDA)**(Lu 2010). This model includes baseline FIQR as part of the response vector while constraining baseline means to be identical across treatment groups, ensuring that differences in FIQR post-randomisation reflect true treatment effects rather than baseline imbalance. This method offers advantages over traditional mixed-effects models (LMM) and ANCOVA by providing a more efficient and unbiased estimate of treatment effects while maintaining the interpretability of a baseline-adjusted analysis(Coffman, Edelman, and Woolson 2016; Lu 2010).

A linear mixed-effects model (LMM) with constraints will be used to compare FIQR scores between treatment groups across all available time points (baseline, 3-months, 6-months, and 12-months), ensuring that baseline differences are accounted for without requiring explicit adjustment in the fixed effects. The model will include a fixed effect for treatment group, time (categorical), and their interaction to allow for non-parallel treatment effects over time. To account for repeated measures within individuals, a random intercept will be specified for each participant, and an unstructured covariance structure will be assumed for residual errors to allow for flexible correlation between repeated measurements.

The constrained longitudinal model will be specified as:

$$Y_{i,t} = \mu + \gamma_t Time_t + \delta_t(Treatment_i \times Time_t) + \sum_k \beta_k Covariate_{k,i} + u_i + \varepsilon_{i,t}$$

where:

- $Y_{i,t}$ is the FIQR score for participant *i* at time *t*
- μ is the grand mean FIQR score at baseline, constrained to be equal for both treatment groups
- $\gamma_t Time_t$ is the fixed effect of time (categorical), with baseline as the reference category
- δ_t is the interaction effect between treatment and time, estimating the treatment effect at each follow-up time point
- β_k represents covariate adjustment terms for stratification factors (sex, participation in the MRI sub-study)
- $u_i \sim N(0, \sigma_u^2)$ represents the random intercept for participant $\underline{i}$
- $\varepsilon_{i,t} \sim N(0, \sigma_t^2)$ is the residual error term

A random intercept model with an unstructured covariance structure will be used to account for within-subject correlation. Adjusted mean differences at 3 months (primary endpoint), 6 months, and 12 months will be presented with 95% confidence intervals and p-values.

Secondary analysis: As a secondary analysis, the above model will be run including a set of *a priori* baseline covariables which may predict the outcome or missingness. In addition to the stratification factors (sex, participation in the MRI sub-study), these will include age, depression (PHQ-9), and anxiety (GAD-7).

Secondary outcome measures: The same methods outlined above will be used to examine the following secondary outcome measures: subjective cognition (British Columbia Cognitive Complaints Inventory, BC-CCI); objective cognitive performance (performance on the number vigilance task: sensitivity, reaction time, reaction time variability); self-reported sleep quality (insomnia severity index, ISI; Pittsburgh Sleep Quality Inventory, PSQI); observed sleep measures derived from actigraphy (sleep efficiency, sleep fragmentation, total sleep time, wake after sleep onset); and pain (Short Form 36 Bodily Pain Scale, SF36-BPS). All outcome contain measurements at baseline (pre-randomisation), 3-months, 6-months, and 12-months, with the exception of actigraphy and objective cognitive performance which were measured at baseline and 3-months.

Model checking and normalising transformations: Assumptions of normality will be assessed graphically looking at residual and quantile-quantile (QQ) plots, and data transformation will be considered if model assumptions (particularly normality of residuals) are clearly violated. Mixed-effects models are robust to small deviations from normality in the residuals; however, if approximate normality cannot be achieved, data normalising transformations will be considered prior to proceeding to fit the same mixed effects linear model described above on the transformed scale.

Examination of residuals following model fitting will be undertaken including exploration of dependencies with observed covariates. Consideration will be given to augmenting the model to heteroscedastic (level 1) residuals especially when the model is fitted in the original scale, where strong patterns are observed between residual errors and covariates and variance increases with the mean(Goldstein 2005). Comparison of the log-likelihood values of a small number of competing models will be used to decide which model to select (higher log-likelihood values indicate a better fitting model).

6.3 Missing Data

Missing data will be minimised by careful data management. Missing data will be described with reasons given where available; the number and percentage of individuals in the missing category will be presented by treatment arm. No data will be considered spurious in the analysis since all data will be checked and cleaned before analysis.

Missing covariate data: If the amount of missing covariate data leads to exclusion of less than 5% of participants, and it is implausible that such exclusion is associated with outcomes or is not associated with observed outcomes, then missing data will be considered ignorable. Fully adjusted analyses will be based on all randomised participants using imputed covariate data using single conditional imputation. Sensitivity analyses will include adjusted analyses based on those with complete covariate profile.

If the amount of missing covariate data leads to exclusion of more than 5% of participants, and auxiliary variables can be identified that are associated with missingness with predictive value for the missing data, then multiple imputation will be employed, including observed outcomes, randomised treatment, auxiliary covariate data predictive of missingness and all variables which will be used in adjusted analyses. The imputation model will include all participants randomised in the group originally assigned.

Missing outcome data: The primary analysis method proposed is reasonably robust to missing at random (MAR) data conditional on all outcome measurements on all other visits and other covariates included in the model (**Table 1**). To address missing data, a multilevel mixed-effects model (cLDA) will be used, which implicitly accounts for intermittent missing data under the MAR assumption(Lu 2010).

Drop-out indicators will be created and we will report the counts of participants dropping out by arm, as well as the follow-up visit they dropped out of (**Table 3a**). For the purpose of our main analyses reporting outcomes up to 6 months, we consider a participant as having drop-out at a particular visit if all follow-up outcome assessments following that visit are all missing.

We will then examine any association of drop-out with important prognostic factors such as age, sex, education level, depression (PHQ-9), anxiety (GAD-7), and clinically significant SAEs relating to the interventions (Section 6.7), and baseline FIQR scores as well as follow-up FIQR scores observed before drop out.

6.4 Sensitivity Analysis

Complier Average Causal Effect (CACE): A sensitivity analysis will be undertaken to estimate the Complier Average Causal Effect (CACE), which aims to assess the treatment effect among participants who adhere to the assigned intervention(Peugh et al. 2017). CACE estimation accounts for non-compliance by using randomisation as an instrument to identify the causal effect of the treatment among those who would comply if assigned to the intervention.

A complier will be defined as a participant who completes at least 3 of the 6 digital CBT-I sessions, as session #3 contains sleep restriction therapy, thought to be the active component of CBT-I. A two-stage least squares (2SLS) instrumental variable (IV) approach will be employed to estimate the CACE, where randomisation is used as an instrument for treatment compliance. The first stage will model compliance as a function of treatment assignment and baseline covariates associated with compliance and the outcome. The second stage will estimate the effect of the treatment on FIQR, using predicted compliance as an instrumental variable. The exclusion restriction assumption assumes that randomisation only affects FIQR through treatment compliance.

Additionally, a continuous measure of compliance (0–6 sessions completed) will be used to estimate a dose-response relationship between adherence and treatment effect. This will be explored using an instrumental variable approach or a generalised additive model (GAM) adjusting for baseline covariates to account for potential nonlinear associations. The relationship between compliance and outcome will also be visualised to assess potential threshold effects.

Balance in baseline characteristics between compliers and non-compliers will be examined. The robustness of CACE estimates will be explored by conducting sensitivity analyses, including alternative compliance definitions (e.g., different session thresholds) and addressing potential unmeasured confounding using a control function approach.

Informative drop out:

To assess the robustness of results to violations of MAR, Missing Not at Random (MNAR) scenarios will be explored using two complementary approaches:

1. Pattern-Mixture MNAR Model (δ-Based Imputation): A δ-adjusted imputation approach will be applied(Cro et al. 2020), where missing FIQR scores are adjusted by an offset δ, representing the assumption that missing data may be systematically worse (or better) than observed data. Sensitivity parameters will vary as follows:

- Optimistic scenario: δ=+1SD of observed FIQR.
- Neutral scenario: δ=0 (equivalent to MAR).
- Pessimistic scenario: δ=−1 SD of observed FIQR.

This approach will be implemented using the `rctmiss` command in Stata, ensuring that MNAR sensitivity analysis remains consistent with the primary model structure(White, Carpenter, and Horton 2018).

If treatment effects remain stable across MAR and MNAR analyses, findings will be considered robust. If treatment estimates shift significantly under MNAR models, this will indicate potential bias due to informative dropout.

6.5 Pre-specified Subgroup Analysis

We will investigate treatment effect modification by baseline sleep duration, an objective measure derived from actigraphy, to assess whether the effectiveness of digital CBT-I differs based on initial sleep duration. There is evidence suggesting that individuals with insomnia and objectively short sleep duration (<6 hours) may have distinct responses to cognitive-behavioural interventions compared to those with longer sleep durations(Vgontzas et al. 2013; Ballesio et al. 2019). This subgroup analysis aims to explore whether baseline total sleep time (TST) modifies the effect of digital CBT-I on the primary outcome (FIQR). None of the included subgroups are based on post-randomisation patient characteristics or events(Yusuf et al. 1991).

Participants will be categorised into two predefined subgroups:

- **Objective short sleep duration:** Baseline actigraphy-measured TST <6 hours
- **Normal sleep duration**: Baseline actigraphy-measured TST ≥6 hours

A treatment-by-subgroup interaction term will be included in the primary analysis model to formally test for effect modification. The interaction term will be as a continuous variable (TST as a linear predictor) to explore potential dose-response relationships.

To assess the consistency of effects across different levels of baseline sleep duration, estimated treatment differences will be summarised at the 25th, 50th, and 75th percentiles of the TST distribution. Results will be reported with estimated treatment effects and corresponding confidence intervals. Forest plots will be used to visualise subgroup effects.

Subgroup analyses will be conducted on the Analysis of Primary Estimand (ATE) (Treatment Policy Analysis) population. If there are fewer than 15 participants in any subgroup or fewer than 5 participants in one treatment arm within a subgroup, the analysis will not be conducted. Given the exploratory nature of this analysis results will be interpreted cautiously.

6.6 Supplementary/ Additional Analyses and Outcomes

In addition to the primary and secondary analyses, the following supplementary neuroimaging analyses will be conducted to provide further insight into the effects of digital CBT-I. These analyses focus on neuroimaging outcomes and their relationship with treatment response. Given their exploratory nature, results will be interpreted cautiously.

Magnetic Resonance Spectroscopy (MRS): Changes in neurotransmitter concentrations in the posterior insula at 3 months post-randomisation will be assessed using MRS. The primary focus will be on glutamate and γ-aminobutyric acid (GABA) levels, given their roles in pain processing and sleep regulation. Between-group differences will be examined using analysis of covariance (ANCOVA), adjusting for baseline neurotransmitter levels.

Resting-State Functional MRI (fMRI): Resting-state functional MRI will be used to examine changes in neuromarkers of sleep quality, pain, and sustained attention at 12 weeks. Functional connectivity within key attention-related networks, including the default mode network (DMN) and frontoparietal network (FPN), will be assessed. Treatment effects will be evaluated using whole-brain voxel-wise comparisons and seed-based connectivity analysis, adjusting for baseline connectivity measures.

Arterial Spin Labelling (ASL) Perfusion MRI: Changes in regional cerebral blood flow (rCBF) at 3 months will be assessed using ASL. Differences in perfusion within regions of interest associated with insomnia, pain, and

cognitive function (e.g., PAG, amygdala, anterior cingulate cortex, thalamus, and insula) will be examined. Group comparisons will be performed using voxel-wise and region-of-interest analyses, adjusting for baseline perfusion levels.

All neuroimaging analyses will be conducted on participants with usable baseline and follow-up scans, following established quality control procedures, using FSL tools. Sensitivity analyses will explore relationships between neuroimaging changes and clinical outcomes, such as FIQR improvement and changes in sleep parameters.

6.7 Harms

This is a low risk, pragmatic trial where the intervention, dCBT-I, is in common use. Thus, we do not anticipate many SAEs. All adverse events are submitted to the trial investigators only if they arise within 12 months post-randomisation and they fall under the category of an SAE: events resulting in death; life-threatening; required hospitalisation; prolonged hospitalisation; other important medical event which may require medical or surgical intervention to prevent one of the serious outcomes listed. SAEs potentially related and unexpected are recorded on the trial's database and assessed for seriousness, causality and expectedness (**Table 6**).

Table 6. List of serious adverse events

Allocation	**Diagnosis**	**Timing of inset of event**[1]	**Description of event (including signs & symptoms)**	**Action taken to deal with the event (including an treatments)**	**Reason for seriousness** [2]	**Causality: is the event related to the intervention?** [3]	**Expectedness (in relation to what is known about the intervention)** [4]

[1] *post-randomisation weeks*

[2] *Death; Life-threatening; required hospitalisation; prolonged hospitalisation; other important medical event which may require medical or surgical intervention to prevent one of the serious outcomes listed.*

[3] *Possibly related; Probably related; Definitely related*

[4] *Expected; Unexpected*

6.8 Health Economics and Cost Effectiveness (where applicable)

A separate Health Economics Analysis Plan (HEAP) will be written by the trial health economist and all cost effectiveness analysis will be undertaken following that plan by the health economist.

7. SPECIFICATION OF STATISTICAL PACKAGES

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

8. PUBLICATION

This study will has been conducted as part of the portfolio of trials in the NIHR Clinical Research Network at the University of Oxford. It will follow/has followed their Standard Operating Procedures ensuring compliance with the principles of Good Clinical Practice and the Declaration of Helsinki and any applicable regulatory requirements.

9. REFERENCES

Ballesio, A., Mrjv Aquino, S. D. Kyle, F. Ferlazzo, and C. Lombardo. 2019. 'Executive Functions in Insomnia Disorder: A Systematic Review and Exploratory Meta-Analysis', *Front Psychol*, 10: 101.

Bastien, Célyne H, Annie Vallières, and Charles M Morin. 2001. 'Validation of the Insomnia Severity Index as an outcome measure for insomnia research', *Sleep medicine*, 2: 297-307.

Bennett, R. M., A. G. Bushmakin, J. C. Cappelleri, G. Zlateva, and A. B. Sadosky. 2009. 'Minimal clinically important difference in the fibromyalgia impact questionnaire', *J Rheumatol*, 36: 1304-11.

Bennett, R. M., R. Friend, K. D. Jones, R. Ward, B. K. Han, and R. L. Ross. 2009. 'The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties', *Arthritis Res Ther*, 11: R120.

Buysse, Daniel J, Charles F Reynolds III, Timothy H Monk, Susan R Berman, and David J Kupfer. 1989. 'The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research', *Psychiatry research*, 28: 193-213.

Carney, C. E., D. J. Buysse, S. Ancoli-Israel, J. D. Edinger, A. D. Krystal, K. L. Lichstein, and C. M. Morin. 2012. 'The consensus sleep diary: standardizing prospective sleep self-monitoring', *Sleep*, 35: 287-302.

Chen, P. Y., Y. W. Jan, and C. M. Yang. 2017. 'Are the Insomnia Severity Index and Pittsburgh Sleep Quality Index valid outcome measures for Cognitive Behavioral Therapy for Insomnia? Inquiry from the perspective of response shifts and longitudinal measurement invariance in their Chinese versions', *Sleep Med*, 35: 35-40.

Choy, Ernest H. S. 2015. 'The role of sleep in pain and fibromyalgia', *Nature Reviews Rheumatology*, 11: 513-20.

Clauw, D. J. 2014. 'Fibromyalgia: a clinical review', *Jama*, 311: 1547-55.

Coffman, Cynthia J, David Edelman, and Robert F Woolson. 2016. 'To condition or not condition? Analysing 'change' in longitudinal randomised controlled trials', *BMJ Open*, 6: e013096.

Crawford, Megan R, Annemarie I Luik, Colin A Espie, Hannah L Taylor, Helen J Burgess, Alex L Jones, Rush University Sleep Research Team, and Jason C Ong. 2020. 'Digital cognitive behavioral therapy for insomnia in women with chronic migraines', *Headache: the journal of head and face pain*, 60: 902-15.

Cro, Suzie, Tim P Morris, Michael G Kenward, and James R Carpenter. 2020. 'Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide', *Statistics in Medicine*, 39: 2815-42.

Diggle, Peter, and Michael G Kenward. 1994. 'Informative drop-out in longitudinal data analysis', *Journal of the Royal Statistical Society Series C: Applied Statistics*, 43: 49-73.

Fayaz, A, P Croft, RM Langford, LJ Donaldson, and GT Jones. 2016. 'Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies', *BMJ Open*, 6: e010364.

Fitzcharles, Mary-Ann, Mary Brachaniec, Lynn Cooper, Ruth Dubin, Trudy Flynn, Kerstin Gerhold, Winfried Häuser, Katherine Cowan, Andreas Laupacis, Renee Marleau, Marc Milot, Nicole Szajcz-Keller, Janice Sumpton, Zach Walsh, and Hani El-Gabalawy. 2017. 'A paradigm change to inform fibromyalgia research priorities by engaging patients and health care professionals', *Canadian Journal of Pain*, 1: 137-47.

Framework, I Conceptual. 1992. 'The MOS 36-item short-form health survey (SF-36)', *Med Care*, 30: 473-83.

Gamble, C., A. Krishan, D. Stocken, S. Lewis, E. Juszczak, C. Doré, P. R. Williamson, D. G. Altman, A. Montgomery, P. Lim, J. Berlin, S. Senn, S. Day, Y. Barbachano, and E. Loder. 2017. 'Guidelines for the Content of Statistical Analysis Plans in Clinical Trials', *Jama*, 318: 2337-43.

Goldstein, Harvey. 2005. 'Heteroscedasticity and complex variation', *Encyclopedia of statistics in behavioral science*, 2: 790-95.

Iverson, G. L., and R. W. Lam. 2013. 'Rapid screening for perceived cognitive impairment in major depressive disorder', *Ann Clin Psychiatry*, 25: 135-40.

Kim, Scott YH, James Flory, and Clare Relton. 2018. 'Ethics and practice of Trials within Cohorts: An emerging pragmatic trial design', *Clinical Trials*, 15: 9-16.

Kyle, Simon D, Madeleine ED Hurry, Richard Emsley, Antonia Marsden, Ximena Omlin, Amender Juss, Kai Spiegelhalder, Lampros Bisdounis, Annemarie I Luik, and Colin A Espie. 2020. 'The effects of digital cognitive behavioral therapy for insomnia on cognitive function: a randomized controlled trial', *Sleep*, 43: 1-12.

Lu, K. 2010. 'On efficiency of constrained longitudinal data analysis versus longitudinal analysis of covariance', *Biometrics*, 66: 891-6.

'National Institute for Health and Care Excellence: Guidelines.' in. 2021. *Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain* (National Institute for Health and Care Excellence (NICE)
Copyright © NICE 2021.: London).

Peugh, James L., Daniel Strotman, Meghan McGrady, Joseph Rausch, and Susmita Kashikar-Zuck. 2017. 'Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer', *Journal of School Psychology*, 60: 7-24.

Riemann, Dieter, Colin A. Espie, Ellemarije Altena, Erna Sif Arnardottir, Chiara Baglioni, Claudio L. A. Bassetti, Celyne Bastien, Natalija Berzina, Bjørn Bjorvatn, Dimitris Dikeos, Leja Dolenc Groselj, Jason G. Ellis, Diego Garcia-Borreguero, Pierre A. Geoffroy, Michaela Gjerstad, Marta Gonçalves, Elisabeth Hertenstein, Kerstin Hoedlmoser, Tuuliki Hion, Brigitte Holzinger, Karolina Janku, Markus Jansson-Fröjmark, Heli Järnefelt, Susanna Jernelöv, Poul Jørgen Jennum, Samson Khachatryan, Lukas Krone, Simon D. Kyle, Jaap Lancee, Damien Leger, Adrian Lupusor, Daniel Ruivo Marques, Christoph Nissen, Laura Palagini, Tiina Paunio, Lampros Perogamvros, Dirk Pevernagie, Manuel Schabus, Tamar Shochat, Andras Szentkiralyi, Eus Van Someren, Annemieke van Straten, Adam Wichniak, Johan Verbraecken, and Kai Spiegelhalder. 2023. 'The European Insomnia Guideline: An update on the diagnosis and treatment of insomnia 2023', *Journal of sleep research*, 32: e14035.

Salazar-Méndez, Joaquín, Nelson Viscay-Sanhueza, Catalina Pinto-Vera, Felipe Oyarce-Contreras, María-Francisca Parra-Vera, Luis Suso-Martí, Eduardo Guzmán-Muñoz, Rubén López-Bueno, Rodrigo Núñez-Cortés, and Joaquín Calatayud. 2024. 'Cognitive behavioral therapy for insomnia in people with chronic musculoskeletal pain. A systematic review and dose-response meta-analysis', *Sleep medicine*, 122: 20-26.

Sarzi-Puttini, Piercarlo, Valeria Giorgi, Daniela Marotto, and Fabiola Atzeni. 2020. 'Fibromyalgia: an update on clinical characteristics, aetiopathogenesis and treatment', *Nature Reviews Rheumatology*, 16: 645-60.

Schulz, Kenneth F., and David A. Grimes. 2005. 'Multiplicity in randomised trials I: endpoints and treatments', *The Lancet*, 365: 1591-95.

Soni, Anushka, Stephanie Santos-Paulo, Andrew Segerdahl, M Kassim Javaid, Rafael Pinedo-Villanueva, and Irene Tracey. 2020. 'Hospitalization in fibromyalgia: a cohort-level observational study of in-patient procedures, costs and geographical variation in England', *Rheumatology*, 59: 2074-84.

Spaeth, Michael. 2009. 'Epidemiology, costs, and the economic burden of fibromyalgia', *Arthritis research & therapy*, 11: 117.

Tang, N. K., S. T. Lereya, H. Boulton, M. A. Miller, D. Wolke, and F. P. Cappuccio. 2015. 'Nonpharmacological Treatments of Insomnia for Long-Term Painful Conditions: A Systematic Review and Meta-analysis of Patient-Reported Outcomes in Randomized Controlled Trials', *Sleep*, 38: 1751-64.

Vgontzas, A. N., J. Fernandez-Mendoza, D. Liao, and E. O. Bixler. 2013. 'Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder', *Sleep Med Rev*, 17: 241-54.

Whibley, Daniel, Kevin Stelfox, Alasdair L Henry, Nicole KY Tang, and Anna L Kratz. 2022. 'Development of a hybrid sleep and physical activity improvement intervention for adults with osteoarthritis-related pain and sleep disturbance: a focus group study with potential users', *British Journal of Pain*, 16: 136-48.

White, I. R., J. Carpenter, and N. J. Horton. 2018. 'A mean score method for sensitivity analysis to departures from the missing at random assumption in randomised trials', *Stat Sin*, 28: 1985-2003.

Wolfe, F., D. J. Clauw, M. A. Fitzcharles, D. L. Goldenberg, W. Hauser, R. L. Katz, P. J. Mease, A. S. Russell, I. J. Russell, and B. Walitt. 2016. '2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria', *Semin Arthritis Rheum*, 46: 319-29.

Yusuf, Salim, Janet Wittes, Jeffrey Probstfield, and Herman A Tyroler. 1991. 'Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials', *Jama*, 266: 93-98.

Zachariae, R., M. S. Lyby, L. M. Ritterband, and M. S. O'Toole. 2016. 'Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials', *Sleep Med Rev*, 30: 1-10.

Zhao, Sijia, Kengo Shibata, Peter J. Hellyer, William Trender, Sanjay Manohar, Adam Hampshire, and Masud Husain. 2022. 'Rapid vigilance and episodic memory decrements in COVID-19 survivors', *Brain Communications*, 4.

APPENDIX: GLOSSARY OF ABBREVIATIONS

CI	Chief Investigator
DSMC	Data and Safety Monitoring Committee
HEAP	Health Economic Analysis Plan
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SUSAR	Serious Unexpected Adverse Reaction
TSC	Trial Steering Committee