STATISTICAL ANALYSIS PLAN

Low Dose Naltrexone for the Treatment of Fibromyalgia:

Statistical Analysis Plan for the Randomized Placebo-Controlled FINAL Trial

Karin Due Bruun[1,2,3], Kirstine Amris[3,4], Henrik Bjarke Vaegter[1,2], Morten Rune Blichfeldt-Eckhardt[1,2], Anders Holsgaard-Larsen[2,5], Robin Christensen[2,3], Palle Toft[2,6]

Affiliations

[1] *Pain Research Group, Pain Center, Odense University Hospital, Odense, Denmark*

[2] *Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark*

[3] *The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark*

[4] *Department of Rheumatology, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark.*

[5] *Department of Orthopedics and Traumatology, Odense University Hospital, Odense, Denmark*

[6] *Department of Anesthesiology and Intensive Care, Odense University Hospital, Odense, Denmark*

Correspondence to: Karin Due Bruun, Pain Research Group, Pain Center, Odense University Hospital, Denmark; karin.due.bruun@rsyd.dk

SAP: Version 1.0. Date 18.05.2022.

Protocol: Version 5.1. Date 27.07.2021.

Roles and responsibility

Authors of the SAP:	Karin Due Bruun and Robin Christensen
Statistician responsible:	Robin Christensen, BSc, MSc, Ph.D.
Primary investigator:	Karin Due Bruun, MD, Ph.D. fellow.
Sponsor:	Palle Toft, MD, Ph.D. Med.Sc.D.

Summary

Objectives:

The primary aim of this randomized placebo-controlled trial is to investigate if 12 weeks of LDN treatment is superior to placebo in reducing pain in women with fibromyalgia. The primary objective is to compare the effect of LDN treatment, relative to placebo, on changes in average pain intensity from baseline to week 12, in women with fibromyalgia. Key secondary outcomes (also assessed after 12 weeks) are other fibromyalgia-related symptoms, i.e., tenderness, fatigue, sleep disturbance, stiffness, memory problems, depression, anxiety as well as global assessment measures, physical function, fibromyalgia impact, pain distribution, and health-related quality of life.

Methods:

A single-center, permuted block randomized, double-blind, placebo-controlled, parallel-group trial will be performed in Denmark. Randomization (1:1) comprises a total of 100 women aged 18-64 years and diagnosed with fibromyalgia who will be treated with either LDN or placebo for 12 weeks, including a four-week titration phase. Our primary analyses will estimate between-group differences in the continuous outcomes after 12 weeks for primary and key secondary outcomes; data collection includes multiple, repeated measurements of each patient's outcome to assess differences in outcomes over time. The main analysis will be based on the intention-to-treat population. For continuous data, missing outcome data after baseline will be handled indirectly using Mixed Models for Repeated Measures Data. All *P* values and 95% confidence intervals will be two-sided. We will not apply explicit adjustments for multiplicity. Instead, we will analyze the secondary outcomes using the Hochberg sequential procedure. Secondarily, to ease interpretation, an analysis of responders will be performed, based on participants with more than 15%, 30%, and 50% improvement in pain after 12 weeks for the LDN and placebo groups, using logistic regression analyses.

Trial registration:

EudraCT number 2019-000702-30 (registered on 2019-07-12).

ClincalTrials.gov Identifier: NCT04270877 (registered on 2020-02-17).

Introduction

Background

Fibromyalgia (FM) is a chronic primary pain disorder where generalized pain is present despite no tissue pathology in the painful areas. Instead, these patients show signs of amplification of nociceptive signals or poor inhibitory control of pain, also described as nociplastic pain features (1). The etiology is thought to be multifactorial, including both biological, psychological, and social factors. Physiologic contributors might include activated glia cells with neuroinflammation or allostasis in the endorphin or autonomic nervous system. These possible mechanistic factors might explain FM's wide range of symptoms, including insomnia, fatigue, memory and concentration difficulties, depression, anxiety, and somatic symptoms.

Traditional pharmacological therapies for FM aims at either reducing facilitatory neurotransmitters or increasing inhibitory control. Evidence supports that both antidepressants with noradrenaline reuptake-inhibition and gabapentinoids are effective in reducing FM pain, with numbers needed to treat in the range 3-7. Unfortunately, these treatments are often accompanied by intolerable side effects.

The rationale for this RCT

Low doses of naltrexone (LDN) have shown promising results in a pilot study (2) and one small randomized trial (3) as a possible new effective pharmacological treatment for FM-related pain with fewer side effects. Both studies have investigated the efficacy of 4.5 mg LDN in women with FM, showing a significant improvement in pain compared to placebo and high adherence to the treatment. The pilot study was not blinded, and the RCT had several risk of biases in the design. Hence, larger RCTs with a more robust methodology are warranted to support the findings. Our study group recently conducted a dose-response study showing good tolerability of doses up to 6 mg (4). We estimated the effective dose in 95% of the participants to be 5.4 mg, suggesting that 6 mg might be more beneficial than 4.5 mg to some FM patients.

Therefore, in this RCT, we investigate the efficacy of 6 mg LDN for treating women with fibromyalgia. We have included a titration phase, and if problems should arise concerning tolerability, lower doses than 6 mg are allowed. We will explore the effectiveness of LDN on pain and other key symptoms associated with fibromyalgia.

Objectives

The trial aims to investigate whether treatment with LDN has a superior effect compared to placebo on patient-reported outcomes. The primary objective is to compare the effect of LDN treatment, relative to placebo, on changes in average pain intensity from baseline to week 12, in women with fibromyalgia.

Secondary objectives include evaluating the clinical effectiveness on 21 secondary outcomes covering core symptoms, daily functioning, the impact of FM, quality of life, global impression of change, and derived responder indices.

Methods

Trial design

The study is designed as a single-center, permuted block randomized, double-blind, placebo-controlled trial. Randomization comprises a parallel-group, randomized (1:1) allocation of a total of 100 women aged 18-64 years diagnosed with fibromyalgia, treated with either LDN or placebo for 12 weeks, including a 4-week titration phase (from baseline to week 4).

Framework

The hypothesis testing framework of this trial is to test if LDN is superior to placebo in reducing pain in women with fibromyalgia.

Randomization, allocation concealment, and blinding

A computerized algorithm was generated for randomization by preparing a list of 100 sequential numbers for active or placebo intervention, and randomization was based on permuted blocks of 2–6 individuals. No stratifications were applied to the randomization, and both investigators and outcome assessors were blinded regarding the permuted blocking strategy.

Sample size and power considerations

Using values from our previous dose-response study, we determined that self-reported pain on a 0-10 NRS at baseline had a mean of 6.7 in the target population, with a standard deviation (SD) of 1.5 NRS points.

According to the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) guidelines, a minimal clinical important difference (MCID) is defined as a 15% decrease in pain, corresponding to a reduction of 1.0 NRS points in the present population. Using an MCID of 1.0 NRS, an SD of 1.5, a statistical power of at least 80%, and a statistical significance level of 0.05, 74 patients are required, i.e., 37 patients in each group. Expecting some attrition and drop-out during the 12-week trial period, we decided to include 100 patients (with approximately 50 patients in each group), corresponding to a statistical power of more than 90% to detect a difference between groups in the ITT population.

Statistical interim analyses and stopping guidance

No interim analyses are carried out. Inclusion started in January 2021, and if the intended sample size is not reached 30 months after that, the inclusion of patients will stop at 74 patients, which will ensure a power of 80%.

Timing of final analysis

All primary and key secondary outcomes are analyzed collectively and reported in the primary manuscript. Final analyses will take place after collecting all 12-week follow-up data (expected in January 2023).

Box 1: summarizes the visit schedule for enrolment, interventions, the timing of assessments, and visit "windows."

	STUDY PERIOD						
	Enrolment	**Allocation**	**Post-allocation**				**Follow-up**
Week	***-4-0***	**0**	***2* (tele-phone)***	***4* (tele-phone)***	***8** (tele-phone)***	***12****	***16*****
ENROLMENT:							
Informed consent	X						
Medication history	X	X	X	X	X	X	X
Demographic data	X						
Eligibility screen	X						
Allocation		X					
INTERVENTIONS:							
Low dose naltrexone		♦—	—	—	—	—♦	
Placebo		♦—	—	—	—	—♦	
ASSESSMENTS:							
Vital tests: blood pressure, weight, height		X				X	X
Safety tests: ALAT, Creatinine, GFR, thrombocyte count, bilirubin. ECG	X					X	
hCRP	X						
Blood for biobank	X					X	
PROMs:							
- PHQ-9	X						
- GAD-7	X						
- FIQR		X		X	X	X	X
- PGI-C				X	X	X	X
- EQ-5D		X		X	X	X	X
- EQ-VAS		X		X	X	X	X
PAIN SENSITIVITY:							
-Handheld algometry		X				X	
-Computerized cuff algometry		X				X	
MUSCLE TESTS:							
-Isometric muscle exhaustion of deltoid		X				X	
-30sec stand chair test		X				X	
Compliance assessment						X	
Adverse events		X	X	X	X	X	X

*Visit window +/- 2 days
**Visit window +/- 7 days
ALAT, alanine aminotransferase; GFR, glomerular filtration rate; ECG, Electrocardiogram; PHQ-9, Patient Health Questionnaire – 9 items; GAD-7, Generalized Anxiety Disorder – 7 items; hCRP, high sensitive C-reactive protein; FIQR, Fibromyalgia Impact Questionnaire revised; PGI-C, Patient Global Impression of Change; EQ-5D, EuroQol 5 dimensions; EQ-VAS, EuroQol Visual Analogue Scale.

Statistical principles

Confidence intervals and *P*-values

All 95% confidence intervals and *P* values will be two-sided; i.e., the level of statistical significance is *a priori* defined as a *P*-value < 0.05. We will not apply explicit adjustments for multiplicity; instead, we will analyze and interpret the secondary outcomes using the Hochberg sequential procedure (5). All of the tests (the multiple tests for various outcomes) are performed, and the resultant *P* values are ordered from largest to smallest on a list. If the largest observed *P* value is less than 0.05, all the tests will be considered "statistically significant." Otherwise, if the next largest *P* value is less than 0.05/2 (.025), all the tests except the one with the largest *P* value are considered significant. If not, and the third *P* value in the list is less than 0.05/3 (.017), then all the tests except those with the largest 2 *P* values are considered significant. This will be continued until all the comparisons are made. This approach uses progressively more stringent statistical thresholds (i.e., the most stringent one being the Bonferroni threshold).

Adherence and protocol deviations

The treatment period begins with a 4-week titration phase, starting with a dose of 1.5 mg LDN or placebo, increasing the dose with 1.5 mg every week to 6 mg at week 4. In case of intolerable side effects, delayed increments are allowed. The highest tolerable dose at the end of week 4 defines the maintenance dose for the remaining eight weeks (week 5-12). The per-protocol population is represented by participants with at least 80% adherence to the intervention. Adherence to the intervention is defined as ingestion of at least 80% of the maintenance dose. Adherence is monitored by counting the number of non-ingested tablets. Adherence is calculated as: % adherence = number of tablets ingested week 5-12 / (number of days week 5-12 x maintenance dose/1.5 mg) x 100%. Furthermore, the use of opioids during the intervention period will be considered a protocol violation.

Analysis populations

The main analyses will be based on the Intention to Treat (ITT) population. The ITT principle asserts the effect of a treatment policy (the planned treatment regimen) rather than the actual treatment given (i.e., it is independent of treatment adherence). Accordingly, participants allocated to a treatment group at baseline (X_{LDN} or $X_{Placebo}$) will be followed up, assessed, and analyzed as members of that group, irrespective of their

adherence to the planned course of treatment (i.e., independent of withdrawals). By applying mixed models, we assume that missingness (among the outcomes) is independent of unobserved measurements but dependent on the observed measurements (i.e., assuming data is '*Missing At Random*'). Repeated measures mixed-effects models generate reasonably valid estimates of treatment effects even when based on missing outcome data over time (6).

Trial population

Screening data

Published screening data comprises: the number of days recruiting, the number of people screened, the number of participants included, the number of people screened but not included, and the reason for screen failure.

Eligibility

The following inclusion criteria are applied:

- Women aged 18-64 years
- Understands and writes Danish
- Fulfills the American College of Rheumatology (ACR) 1990 criteria for fibromyalgia
- A minimum score of 4 in self-reported average pain during the last seven days on a 0-10 NRS at baseline
- All fertile women have to use safe anti-conception (spiral, birth control pills, contraceptive patch, contraceptive vaginal ring, or gestagen injections) three weeks before and one week after the trial. If the participants' normal lifestyle includes sexual abstinence, they do not have to use anti-conception. Instead, they can give oral informed consent that they will be sexually abstinent during the trial. A woman is considered non-fertile if she is sterilized, hysterectomized, bilateral oophorectomized, or postmenopausal.

The following exclusion criteria are applied:

- Known allergy against naltrexone hydrochloride
- Pregnancy or breastfeeding. A negative pregnancy test has to be available for all fertile subjects at baseline
- Use of opioids or NSAIDs up to 4 weeks before inclusion in the trial

- Known abuse of alcohol or other substances
- Known inflammatory rheumatic diseases
- Known demyelinating diseases
- Known active cancer
- Liver dysfunction (ALAT must not be elevated more than 2-fold over the highest reference level)
- Kidney dysfunction (GFR < 59 mL/min)
- Psychotic diseases
- History of suicide attempts
- Suicide ideation – evaluated using Patient Health Questionnaire (PHQ) 9 items (item 9 = "never")

Recruitment

The CONSORT flow diagram (Figure 1) will include information about the number of people recruited, screened, eligible, consented, randomized, allocated to treatment, withdrew, and lost to follow-up.

Withdrawal/follow-up

A CONSORT flow diagram (Figure 1) will display the number of participants lost to follow-up for the following time points: before baseline, 4-week visit, 8-week visit, and 12-week visit. The reason for loss to follow-up will be dichotomized into either withdrawal of consent or adverse event. The level of consent withdrawal will be categorized as withdrawal from the intervention, withdrawal from follow-up, or complete withdrawal.

Baseline patient characteristics

Baseline characteristics for each treatment arm will be summarized to visualize whether relevant demographic, pain-related, and other relevant characteristics appear balanced across the two intervention groups. These data will be presented in the primary manuscript as outlined in Table 1. Categorical data will be described using numbers and percentages. Normally distributed continuous data will be described using mean and SD, whereas continuous data that are skewed will be described using median and interquartile range. Baseline characteristics will not be compared for statistically significant differences between the intervention groups. Instead, any clinically relevant differences will be described in the manuscript.

Analysis

Outcome definitions

Primary outcome measure: Change in pain intensity from baseline to after 12 weeks of treatment, using the "level of pain" question from the FIQR (measures the average pain the last seven days on an 11-point rating scale (ranging from 0= "no pain" to 10 = "unbearable pain").

Secondary outcome measures assessing change from baseline after 4, 8, and 12 weeks of treatment, with the 12-week assessment being of primary interest:

- Global assessment: assessed by Patient Global Impression of Change on a 1-7 Verbal Rating Scale.
- Impact of fibromyalgia: assessed by the FIQR total score.
- Pain distribution: assessed by the Widespread Pain Index (WPI) from the 2016 diagnostic criteria for fibromyalgia.
- Level of pain (pain trajectory): assessed by the FIQR "level of pain" question
- Level of tenderness: assessed by the FIQR "level of tenderness to touch" question
- Tenderness - pressure pain threshold: measured using a handheld pressure algometer. Points measured are right quadriceps 15 cm from apex patella and left trapezius 10 cm acromion (between acromion and C6/7). Each point is measured three times; point 2 and 3 is measured 1 cm above and 1 cm below the first point. An average of the six measurements is reported in kiloPascal.
- Level of fatigue: assessed by the FIQR "level of energy" question.
- Level of sleep disturbance: assessed by the FIQR "quality of sleep" question.
- Level of depression: assessed by the FIQR "level of depression" question.
- Level of anxiety: assessed by the FIQR "level of anxiety" question.
- Level of cognition: assessed by the FIQR "level of memory problems" question.
- Level of stiffness: assessed by the FIQR "level of stiffness" question.
- Level of physical function: assessed by the physical function domain of FIQR.
- Health-related quality of life (mobility): assessed by the EQ-5D-5L mobility domain.
- Health-related quality of life (self-care): assessed by the EQ-5D-5L self-care domain.
- Health-related quality of life (usual activities): assessed by the EQ-5D-5L usual activities domain.
- Health-related quality of life (pain/discomfort): assessed by the EQ-5D-5L pain/discomfort domain.

- Health-related quality of life (anxiety/depression): assessed by the EQ-5D-5L anxiety/depression domain.
- Health-related quality of life (global): assessed by the EQ-5D Visual Analogue Scale (EQ-VAS).

Finally, a secondary objective is to investigate the number of responders in both treatment groups.

Three responder categories are defined:

- Number of responders with a more than 15% improvement of the primary outcome.
- Number of responders with a more than 30% improvement of the primary outcome.
- Number of responders with a more than 50% improvement of the primary outcome.

Analysis methods

The pre-specified efficacy analyses will be based on the data for the ITT population, which includes all participants assessed and randomized at baseline. For the primary outcome, the significance level is defined as a p-value < 0.05. Regarding the secondary outcomes, we will not apply explicit adjustments for multiplicity; instead, we will analyze and interpret them based on the Hochberg sequential procedure. Unlike the Bonferroni correction/interpretation (directly adjusting the statistical significance threshold by the number of tests planned [say, k] $\rightarrow$ P^*= 0.05/k), the Hochberg sequential procedure sorts the P-values from largest to smallest on a list. This approach uses progressively more stringent statistical thresholds, with the most stringent one being the Bonferroni threshold. Thus, the approach should achieve a greater power to detect true effect than the Bonferroni procedure under appropriate conditions.

Our primary analyses will estimate between-group differences in the continuous outcomes after 12 weeks of treatment for primary and secondary outcomes. Repeated measurements (T= 0, 4, 8, and 12 weeks from baseline) are used based on a linear mixed-effects model. All between-group differences will be adjusted for baseline level to reduce the random variation. The primary statistical model will consist of fixed effects factors and random effects for patients. Fixed effects factors define the expected values of the observations, and random effects define the variance and covariances of the observations. Participants will be randomly assigned to two treatment groups (X_{LDN} vs. $X_{Placebo}$); observations are made at four time points for the primary outcome measure (baseline and 4, 8, and 12 weeks from baseline). Basically, there are two fixed-effect factors: group and time. Random effects result from variation between and within participants.

We anticipate that measures on the same patient at different times are correlated, with measures taken closely together in time being more highly correlated than measures taken more apart in time. Observations on different participants will be assumed to be independent.

Secondarily, an analysis of number of responders and harms (dichotomous outcomes) in the two groups will be carried out using logistic regression analyses. A responder is defined as a participant who reports a more than 15%, 30%, or 50% decrease in pain after 12 weeks of treatment with LDN. For these dichotomous outcomes, logistic regression will be used to calculate the Odds Ratio (OR) with 95% CI comparing the two groups. For subsequent ease of interpretation, the OR values will be converted into (relative) Risk Ratios and (absolute) Risk Differences.

Missing data

Repeated measurements using mixed-effects models will be based on the ITT population, including all randomized participants with available data at baseline. Missing data will be handled indirectly and statistically modeled using repeated-measures linear mixed models (see below). These models will be valid if data are Missing at Random (MAR): "*Any systematic difference between the missing values and the observed values can be explained by differences in observed data*" (2). Contrasts between groups will be estimated based on least squares means derived from the mixed linear models (i.e., primary contrast at 12 weeks from baseline). The primary analyses for the dichotomous outcomes will be based on a simple, single-step non-responder imputation in the case of any missing binary outcomes after 12 weeks.

Sensitivity analyses will be performed on the main analyses to confirm the robustness of the findings, including the following:

(i) Non-responder imputation: use of a simple single imputation where the baseline observation is carried forward; potentially valuable if data are Not Missing At Random (NMAR)

(ii) 'Per Protocol' population: defined as participants with at least 80% adherence to treatment.

Robustness is a concept that refers to the sensitivity of the overall conclusions to various limitations of the data, assumptions, and analytic approaches to data analysis. Robustness implies that the treatment effect and primary conclusions of the FINAL trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.

Harms

Data on adverse events (AE) and adverse reactions (AR) are collected at all visits. The participants will complete a questionnaire about known side effects. Furthermore, the participants are interviewed about any adverse events occurring during the trial. We will provide information about the causality, expectedness, and severity of the adverse events. PI assesses whether an AE is expected and related to the trial medication using the Summary of Product Characteristics (SmPC) for Naltrexone 50 mg as a reference document. The Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 will be used as the reference document for graduating the severity of harm.

Based on the ITT population, the following safety data will be summarized in the final analysis report:

- The number and percentages of participants who withdraw from the intervention because of adverse events are reported for both treatment arms.
- All adverse events are summarized with a breakdown of severity category (mild, moderate, severe, serious, and death) for both treatment arms, with a further breakdown in doses for the active treatment group (3 mg or below, 4.5 mg, and 6 mg) (Table 3).
- Twelve pre-specified adverse events based on the questionnaire with known side effects will be summarized for both treatment arms, with a further breakdown in doses for the active treatment group (3 mg or below, 4.5 mg, and 6 mg) (Table 3).

Statistical software

All analyses were performed using the statistical software SAS version 9.4 (SAS Institute, Cary, NC, USA).

References

1. Bruun KD, Amris K, Vaegter HB, Blichfeldt-Eckhardt MR, Holsgaard-Larsen A, Christensen R, et al. Low-dose naltrexone for the treatment of fibromyalgia: protocol for a double-blind, randomized, placebo-controlled trial. Trials. 2021;22(1):804.

2. Younger J, Mackey S. Fibromyalgia symptoms are reduced by low-dose naltrexone: a pilot study. Pain Med. 2009;10(4):663-72.

3. Younger J, Noor N, McCue R, Mackey S. Low-dose naltrexone for the treatment of fibromyalgia: findings of a small, randomized, double-blind, placebo-controlled, counterbalanced, crossover trial assessing daily pain levels. Arthritis Rheum. 2013;65(2):529-38.

4. Bruun-Plesner K, Blichfeldt-Eckhardt MR, Vaegter HB, Lauridsen JT, Amris K, Toft P. Low-Dose Naltrexone for the Treatment of Fibromyalgia: Investigation of Dose-Response Relationships. Pain Med. 2020;21(10):2253-61.

5. Cao J, Zhang S. Multiple comparison procedures. JAMA. 2014;312(5):543-4.

6. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. Jama-J Am Med Assoc. 2016;315(4):407-8.

Manuscript Outline

Figure 1. Trial Profile: Patient Flow Throughout the FINAL Trial.

Table 1. Baseline Characteristics for the Intention-To-Treat Population.

Figure 2. The trajectory for the Average Pain (during the last seven days) over time from baseline to endpoint after 12 weeks.

Table 2. Primary, Key Secondary, and Other Secondary Outcomes at 12 Weeks from Baseline

Figure 3. Patient Global Impression of Change (PGI-C) related to the intervention after 12 weeks of treatment, rated on a 7-item Verbal Rating Scale ranging from much worse to much improved.

Table 3. Adverse events in the safety population defined as participants in the ITT population who have received at least one dose of their allocated intervention.

Figure 1. Trial Profile: Patient Flow Throughout the FINAL Trial.

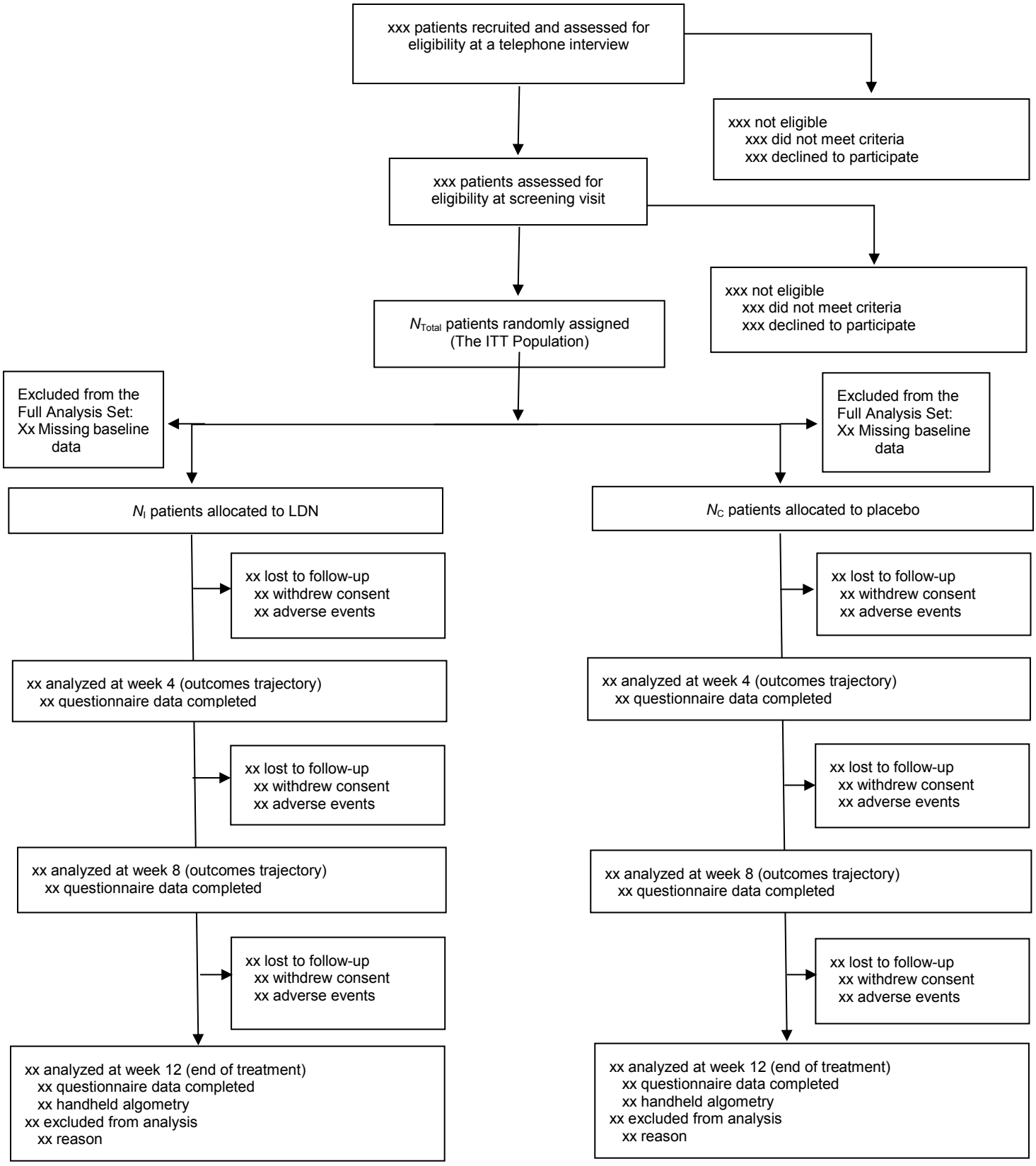

Table 1. Baseline Characteristics for the Intention-To-Treat Population.

	LDN & usual care N_I	Placebo & usual care N_C	Total population N_{Total}
Variables			
Age, years			
Height, meters			
Body Weight, kg			
Body Mass Index, kg/m^2			
Duration of chronic pain, in years			
Pain medication, number (%)			
- no pain medication			
- 1 pain medication			
- 2 or more pain medications			
Concomitant pain medication, number (%):			
- paracetamol			
- tricyclic antidepressants or SNRI[1]			
- gabapentin or pregabalin			
- other			
Outcomes			
Level of pain*, average last seven days (range 0-10)			
FIQR[2] total score (range 0-100)			
Pain distribution, WPI[3] (range 0-19)			
Level of tenderness*, average last 7 days (0-10)			
Tenderness, average pressure pain threshold**, in kPa			
Level of fatigue*, average last 7 days (0-10)			
Level of sleep disturbance*, average last 7 days (0-10)			
Level of depression*, average last 7 days (0-10)			
Level of anxiety*, average last 7 days (0-10)			
Level of cognition*, average last 7 days (0-10)			
Level of stiffness*, average last 7 days (0-10)			
FIQR function domain (range 0-90)			
EQ-5D-5L[4] mobility domain (range 1-5)			
EQ-5D-5L self-care domain (range 1-5)			
EQ-5D-5L usual activities domain (range 1-5)			
EQ-5D-5L pain/discomfort domain (range 1-5)			
EQ-5D-5L anxiety/depression domain (range 1-5)			
EQ-VAS[5] (range 0-100)			

Values will be presented as Means (SDs) unless otherwise stated. [1]Serotonin-Noradrenalin-Reuptake-inhibitor, [2]Fibromyalgia Impact Questionnaire-Revised, [3]Widespread Pain Index, [4]EuroQoL 5 dimensions 5 levels, [5]EuroQoL Visual Analog Scale

*FIQR item

**Measured using a handheld pressure algometer, points measured are right quadriceps 15 cm from apex patella and left trapezius 10 cm acromion (between acromion and C6/7). Each point is measured three times; point 2 and 3 is measured 1 cm above and 1 cm below the first point. An average of the six measurements is reported.

Figure 2. The trajectory for the Average Pain (during the last seven days) over time from baseline to endpoint after 12 weeks

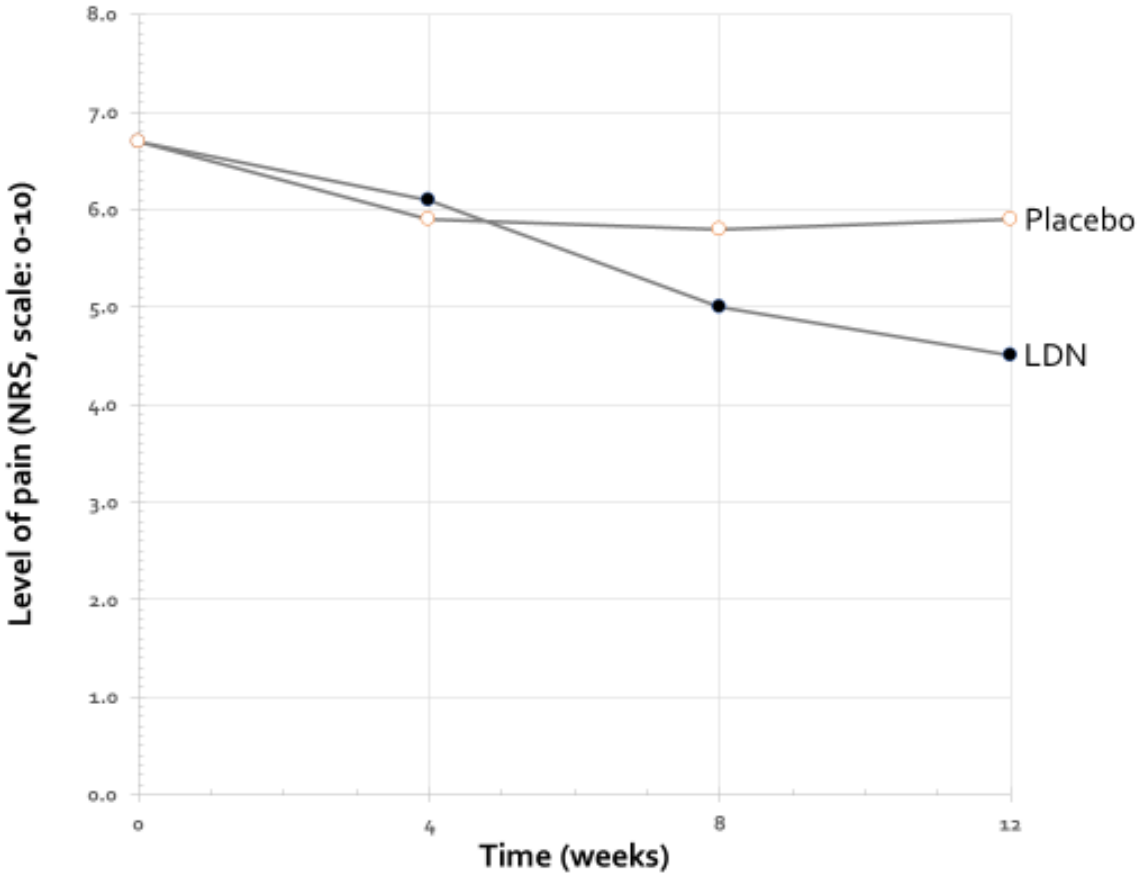

Table 2. Primary, Key Secondary, and Other Secondary Outcomes at 12 Weeks from Baseline based on the ITT population (Repeated Measures Mixed Effects Models)

	Change from baseline to after 12 weeks of treatment (95% CI)		Difference between groups (95% CI)	*P*-value
	LDN	**Placebo**		
Primary Outcome: Level of pain* NRS[1] 0-10				
Key Secondary Outcomes:				
Global impression of change				
Impact of fibromyalgia, FIQR[2] total score 0-100				
Pain distribution, WPI[3] 0-19				
Level of tenderness* NRS 0-10				
Average pain pressure threshold**				
Level of fatigue* NRS 0-10				
Level of sleep disturbance* NRS 0-10				
Level of depression* NRS 0-10				
Level of anxiety* NRS 0-10				
Level of cognition* NRS 0-10				
Level of stiffness* NRS 0-10				
Level of physical function, FIQR function domain 0-90				
Health-related quality of life (mobility) EQ-5D-5L[4] domain, VRS 1-5				
Health-related quality of life (self-care) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (usual activities) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (pain/discomfort) EQ-5D-5L domain VRS 1-5				
Health-related quality of life (anxiety/depression) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (global) EQ-VAS[5] 0-100				
[α]Responder indices:				
15% improvement in pain, number (%)				
30% improvement in pain, number (%)				
50% improvement in pain, number (%)				

Estimates will be presented as Least Squares Means (and Standards Errors) per group, and the difference between them will be reported with the corresponding 95% Confidence Interval – unless otherwise stated.
[1]Numeric Rating Scale, [2]Fibromyalgia Impact Questionnaire-Revised, [3]Widespread Pain Index, [4]EuroQoL 5 dimensions 5 levels, [5]EuroQoL Visual Analog Scale
*Items from the FIQR
**Measured using a handheld pressure algometer, points measured are right quadriceps 15 cm from apex patella and left trapezius 10 cm acromion (between acromion and C6/7). Each point is measured three times; point 2 and 3 is measured 1 cm above and 1 cm below the first point. An average of the six measurements is reported in kilopascal.
[α]Missing data will be replaced using a (single-step) non-responder imputation.

Figure 3. Patient Global Impression of Change (PGI-C) related to the intervention after 12 weeks of treatment, rated on a 7-item Verbal Rating Scale ranging from much worse to much improved. Data will be analyzed based on the ITT population, while only data as observed will be applied (i.e., respecting the original randomization).

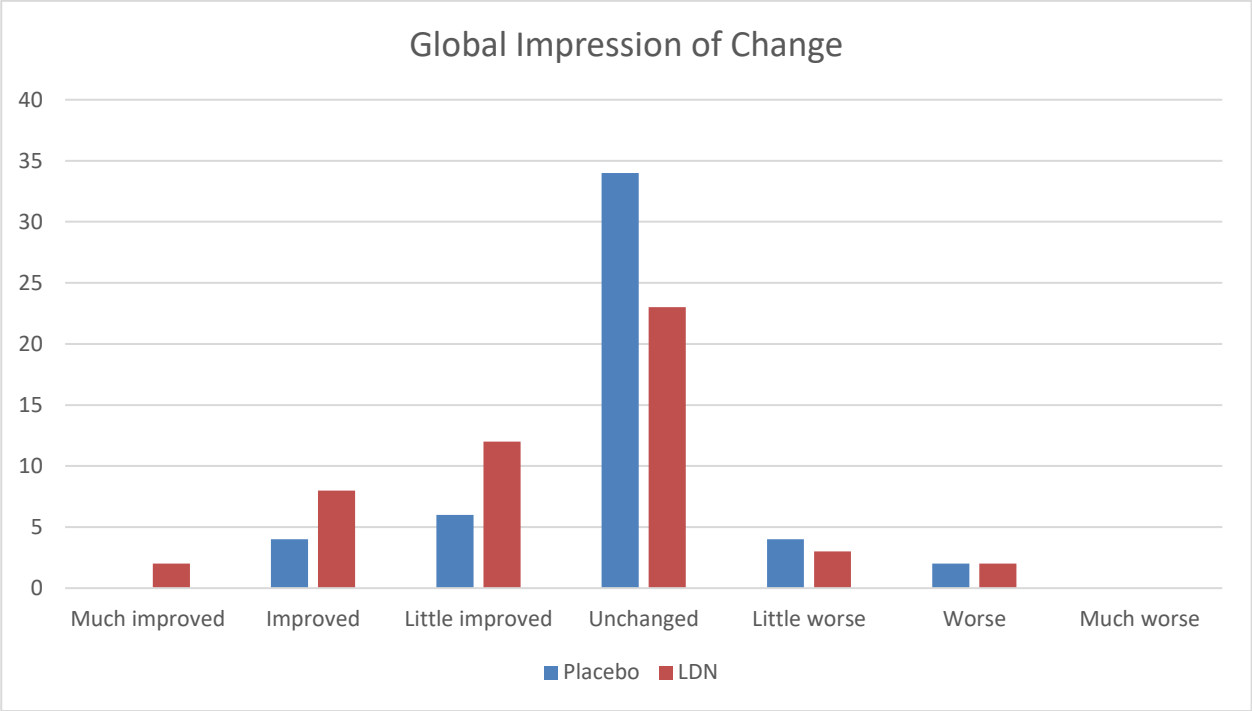

Table 3. Adverse events in the safety population defined as participants in the ITT population who have received at least one dose of their allocated intervention.

	LDN			Placebo
	1.5-3 mg	**4.5 mg**	**6 mg**	
Exposure time (patient weeks)				
AE, n patients (%)				
AE, n events (rate per patient week)				
AE mild, n patients (%)				
AE moderate, n patients (%)				
AE severe, n patients (%)				
SAE, n patients (%)				
Deaths, n patients (%)				
Pre-specified adverse events				
Headache, n patients (%)				
Vivid dreams, n patients (%)				
Diarrhea, n patients (%)				
Constipation, n patients (%)				
Abdominal ache, n patients (%)				
Nausea, n patients (%)				
Increased appetite, n patients (%)				
Dizziness, n patients (%)				
Palpitations, n patients (%)				
Hot flashes, n patients (%)				
Dry mouth, n patients (%)				
Depressed mood, n patients (%)				

AE, adverse event
SAE, serious adverse event.
The severity of an AE refers to the maximum intensity of the event. An event is considered mild if it does not interfere with activities of daily life (ADL), moderate if it limits instrumental ADL, and severe if it interferes substantially with the patient's ADL. An AE is classified as serious if fatal or life-threatening, requires inpatient hospitalization, causes significant disabling, or requires medical intervention to prevent permanent impairment or damage.

Pre-specified Sensitivity and Ancillary Analyses

Appendix Table 1. Primary and Key Secondary Outcomes at 12 Weeks from Baseline, based on the ITT population where missing data is replaced using a single-step non-responder imputation technique for continuous outcomes.

	Change from baseline to after 12 weeks of treatment (95% CI)		Difference between groups (95% CI)	*P*-value
	LDN	**Placebo**		
Primary Outcome: Level of pain* NRS[1] 0-10				
Key Secondary Patient Reported Outcomes Global impression of change				
Impact of fibromyalgia, FIQR[2] total score 0-100				
Pain distribution, WPI[3] 0-19				
Level of tenderness* NRS 0-10				
Average pain pressure threshold**				
Level of fatigue* NRS 0-10				
Level of sleep disturbance* NRS 0-10				
Level of depression* NRS 0-10				
Level of anxiety* NRS 0-10				
Level of cognition* NRS 0-10				
Level of stiffness* NRS 0-10				
Level of physical function, FIQR function domain 0-90				
Health-related quality of life (mobility) EQ-5D-5L[4] domain, VRS 1-5				
Health-related quality of life (self-care) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (usual activities) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (pain/discomfort) EQ-5D-5L domain VRS 1-5				
Health-related quality of life (anxiety/depression) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (global) EQ-VAS[5] 0-100				

Estimates will be presented as Least Squares Means (and Standards Errors) per group, and the difference between them will be reported with the corresponding 95% Confidence Interval – unless otherwise stated.
[1]Numeric Rating Scale, [2]Fibromyalgia Impact Questionnaire-Revised, [3]Widespread Pain Index, [4]EuroQoL 5 dimensions 5 levels, [5]EuroQoL Visual Analog Scale
*Items from the FIQR
**Measured using a handheld pressure algometer, points measured are right quadriceps 15 cm from apex patella and left trapezius 10 cm acromion (between acromion and C6/7). Each point is measured three times; point 2 and 3 is measured 1 cm above and 1 cm below the first point. An average of the six measurements is reported in kilopascal.

Appendix Table 2. Primary and Key Secondary Outcomes at 12 Weeks from Baseline, based on the Per Protocol population (Repeated Measures Mixed Effects Models).

	Change from baseline to after 12 weeks of treatment (95% CI)		Difference between groups (95% CI)	*P*-value
	LDN	**Placebo**		
Primary Outcome: Level of pain* NRS[1] 0-10				
Key Secondary Patient Reported Outcomes Global impression of change				
Impact of fibromyalgia, FIQR[2] total score 0-100				
Pain distribution, WPI[3] 0-19				
Level of tenderness* NRS 0-10				
Average pain pressure threshold**				
Level of fatigue* NRS 0-10				
Level of sleep disturbance* NRS 0-10				
Level of depression* NRS 0-10				
Level of anxiety* NRS 0-10				
Level of cognition* NRS 0-10				
Level of stiffness* NRS 0-10				
Level of physical function, FIQR function domain 0-90				
Health-related quality of life (mobility) EQ-5D-5L[4] domain, VRS 1-5				
Health-related quality of life (self-care) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (usual activities) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (pain/discomfort) EQ-5D-5L domain VRS 1-5				
Health-related quality of life (anxiety/depression) EQ-5D-5L domain, VRS 1-5				
Health-related quality of life (global) EQ-VAS[5] 0-100				

Estimates will be presented as Least Squares Means (and Standards Errors) per group, and the difference between them will be reported with the corresponding 95% Confidence Interval – unless otherwise stated.
[1]Numeric Rating Scale, [2]Fibromyalgia Impact Questionnaire-Revised, [3]Widespread Pain Index, [4]EuroQoL 5 dimensions 5 levels, [5]EuroQoL Visual Analog Scale
*Items from the FIQR
**Measured using a handheld pressure algometer, points measured are right quadriceps 15 cm from apex patella and left trapezius 10 cm acromion (between acromion and C6/7). Each point is measured three times; point 2 and 3 is measured 1 cm above and 1 cm below the first point. An average of the six measurements is reported in kilopascal.