Document Type:	Statistical Analysis Plan
Official Title:	A randomized, parallel-group treatment, Phase 2, double-blind, pilot study to investigate the efficacy and safety of elinzanetant compared with placebo for treatment of sleep disturbances associated with menopause
NCT Number:	NCT06112756
Document Date:	15 NOV 2024

Title Page

Protocol Title:

A randomized, parallel-group treatment, Phase 2, double-blind, pilot study to investigate the efficacy and safety of elinzanetant compared with placebo for treatment of sleep disturbances associated with menopause.

Protocol Number:	22423
Compound Number:	BAY 3427080 / elinzanetant
Short Title:	A pilot study to assess the effects of elinzanetant on sleep disturbances associated with menopause.
Acronym:	NIRVANA
Sponsor Name:	Bayer Consumer Care AG
Legal Registered Address:	Peter-Merian-Strasse 84, 4052 Basel, Switzerland

Regulatory Agency Identifier Number(s):

Registry	ID
EU CT	2023-504955-28-00

Date: 15 NOV 2024

Version: 2.0

Statistical Analysis Plan template version: 4.1

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.
The approval of the Statistical Analysis Plan is documented in a separate signature document.

Table of Contents

Table of Tables

Table of Figures

Version History

This Statistical Analysis Plan (SAP) for Study 22423 is based on the protocol Version 2.1 dated 06 FEB 2024.

SAP Version	Date	Change	Rationale
1.0	17 JAN 2024	Not applicable	Original version
2.0	13 NOV 2024	This amendment of the SAP includes corrections of typos, clarifications of wording and analyses, and additions of data handling rules and analyses. Major changes have been incorporated into the following sections:	Changes were identified during review of this document and of blinded data.
		1) Clarified definitions and additional details of data handling when identifying intercurrent events (Section 1.1 and Section 6.4 Appendix 4: identification of intercurrent events)	Clarifications made based on blinded data programming.
		2) Added 95% confidence interval to be summarized for descriptive statistics and 90% confidence interval from model analysis (Section 4.1 General Considerations).	Changes made based on TFL specification and blinded review of tables and figures.
		3) Clarified and modified data rules (Section 4.1.2 Data Rules)	Changes made based on blinded data programming.
		4) Removed individual profile plots of intercurrent event (Section 4.1.3 Intercurrent Events)	Changes made based on blinded data programming.
		5) Modified visit windows for unscheduled and end of treatment PSG assessments (Section 4.2.1 Definition of Endpoints).	Changes made based on blinded data review.
		6) Added descriptive statistics of WASO and sleep efficiency using Sleepiz One+ analysis set (Section 4.2.2 and Section 4.3.1.2)	Change made based on blinded review of tables and figures.

		7) Clarified PROC MI model includes indicator variables for stratification variable (Section 4.2.2.1)	Modifications needed based on blinded data programming.
		8) Clarified details of the supplemental analyses and added scatterplot and QQ plots for supplementary 2 analysis (Section 4.2.3 Supplementary Analyses).	Clarifications needed based on blinded data review.
		9) For PROs assessments, mapping rules to study week were clarified, and data handling rules for baseline derivation were modified (Section 4.3.2, Section 4.4.2, Section 4.4.3, Section 4.4.4, Section 4.4.5, Section 4.6.3)	Modifications needed based on blinded data programming.
		10) Data handling rules added to exclude sleep diary data that do not logically make sense from analyses (Section 4.4.2).	Rules added based on blinded data review.
		11) Added rule to impute missing 2nd night 'Study medication intake documentation' data at week 12 if eCRF data indicate that subject received treatment for calculation of treatment compliance using eDiary data (Section 4.5.1).	Rules added based on blinded data review.
		12) Additional analyses on hot flash daily diary data were added (Section 4.4.3)	Changes made based on TFL specification and blinded review of tables and figures.
		13) Individual participant presentations in Section 4.5.3.2 were modified	Figures and Listing were modified to be consistent with advice from OASIS pre-NDA meeting with FDA.
		14) Modified to take average value if more than 1 measurement is available on the same date for vital signs (Section 4.5.3.3).	Rules added based on blinded data review.

		15) For Sleepiz One+ data, added if recording started on or before 11am then date of observation will be the day before, added rules to exclude data recorded after end of treatment, and added analysis using only nights with GSD ≥ 4 hours (Section 4.6.3)	Rules added based on blinded data review.
		16) Modified summary of hysterectomy to summary of oophorectomy (Section 6.1.3)	Changes made based blinded review of tables and figures.
		17) Added summary of relevant prior and concomitant procedures (Section 6.1.6)	Added based on blinded data review.
		18) Modified definition of average awakening from Sleepiz One+ data (Section 6.2.2)	Modified to reflect the data based on blinded data review.
		19) Preliminary list of prohibited therapies was added. Two tables were added to allow specification of prohibited hormonal therapies and prohibited concomitant medication impacting PK by drug groupings, and one table to allow specification of prohibited concomitant medications potentially confounding efficacy by drug names (Section 6.3).	Added based on blinded data review.
		20) For AESI “Post-menopausal bleeding” typo corrected, deleted duplicate term and added new term “PT Unexpected vaginal bleeding on hormonal IUD” (Section 6.5).	Update of applicable coding conditions.

List of Abbreviations and Definitions of Terms

AE(s)	Adverse event(s)
AESI	Adverse event of special interest
AHI	Apnea hypopnea index / Respiratory Irregularity Index (for Sleepiz One+)
ALT	Alanine aminotransferase
AP	Alkaline phosphatase
AST	Aspartate aminotransferase
BMI	Body mass index
CK	Creatinine kinase
CV	Coefficient of variation
eCRF(s)	Electronic case report form(s)
eC-SSRS	Electronic Columbia-Suicide Severity Rating Scale
eDiary	Electonic diary
EMWT	Early morning wake time
EoT	End of treatment
ePRO(s)	Electronic patient-reported outcome(s)
GAD-7	General Anxiety Disorder-7
GGT	Gamma-glutamyl
HF(s)	Hot flash(es)
HFDD	Hot Flash Daily Diary
ICE(s)	Intercurrent event(s)
ICF	Informed consent form
INR	International normalized ratio
IRT	Interactive response technology
ISI	Insomnia Severity Index
LDH	Lactate dehydrogenase
LLT	Lower Level Term
LPS	Latency to persistent sleep
LSM	Least square means
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Affairs
MMRM	Mixed model with repeated measures
NAW	Number of awakenings
NK / NK-1 / NK-3	Neurokinin / neurokinin 1 / neurokinin 3
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PHQ-9	Patient Health Questionnaire-9
PK	Pharmacokinetic(s)
PLMAI	Periodic limb movement disorders with arousal index
PROMIS SD SF 8b	Patient-Reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b
PSG	Polysomnography
PT	Preferred term
REM	Rapid eye movement
SAE(s)	Serious adverse event(s)
SAP	Statistical analysis plan
SAS	Statistical analysis software
SCR-1/ SCR-2	Screening Visit 1 / screening Visit 2
SD	Sleep Diary
SE	Sleep efficiency
SLAS	Sleepiz One+ analysis set
SoA	Schedule of assessment
SOC	System organ class
SOL	Sleep onset latency
SOLN1	Sleep onset latency to N1
sLPS	Subject-reported latency to persistent sleep
sNAW	Subject-reported number of awakenings

sTIB	Subject-reported time in bed
sTST	Subject-reported total sleep time
sSE	Subject-reported sleep efficiency
sWASO	Subject-reported wakefulness after sleep onset
T1/T2/T3	Treatment Visit 1/ Treatment Visit 2/ Treatment Visit 3
TIB	Time in bed
TB	Total bilirubin
TEAE(s)	Treatment-emergent adverse event(s)
TST	Total sleep time
TWT	Total wake time
ULN	Upper limit of normal
VMS	Vasomotor symptoms
WASO	Wakefulness after sleep onset
WTDS	Wake time during sleep

1. Introduction

Menopause is a stage of natural aging that marks the end of a woman's reproductive period and is characterized by various physiological changes. Among menopausal symptoms, sleep disturbances are one of the most debilitating symptoms that occur during menopause. Nighttime awakenings due to VMS (also known as night-sweats or hot flashes) are one instigator of sleep disturbances, however, sleep disturbances associated with menopause are not completely explained by nocturnal vasomotor symptoms and likely have a multifactorial basis (Baker et al. 2018a, Woods et al, 2016). The effect of hormone treatment on sleep disturbances in this group of women is small (Guthrie et al. 2018) indicating that additional biological mechanisms beyond reduced estrogen receptor signaling and nighttime awakening due to VMS may contribute to sleep disturbances during the menopausal period.

Clinical studies using objective measurement methods have identified increases in nighttime awakenings and WASO as key aspects of sleep disturbance in peri-/postmenopausal women but have not found differences in total sleep time or hyperarousal compared with younger women (Baker et al. 2018b, Coborn et al. 2022, Pengo et al. 2018).

A small double-blind, randomized cross-over study in healthy male volunteers confirmed the involvement of Substance P in sleep disturbances. After intravenous infusion of Substance P both polysomnography and subjective rating scales indicated a significant decrease of sleep quantity and quality (Lieb et al. 2002). Antagonizing Substance P activity with a NK-1 receptor antagonist is therefore expected to show an improvement in sleep. A much larger study has shown that vestipitant, a NK-1 receptor antagonist improves WASO in 161 male and female patients with primary insomnia (Ratti et al. 2013).

Elinzanetant (formerly NT-814[1]) is a dual neurokinin (NK)-1,3 receptor antagonist and is in development for the treatment of moderate to severe vasomotor symptoms. SWITCH-1 (814-PM-02), a Phase 2b study, showed statistically significant and clinically meaningful improvements on elinzanetant 120 mg and 160 mg once daily compared to placebo in global Pittsburgh Sleep Quality Index (PSQI) and total Insomnia Severity Index (ISI) (Simon et al. 2023).

If benefits of elinzanetant on sleep are confirmed in further studies by e.g. subjective and objective measures of sleep quality, it is anticipated that elinzanetant will be a relevant improvement for women with sleep disturbances associated with menopause.

Study 22423 is a phase 2 clinical trial to assess the efficacy and safety of elinzanetant for the treatment of sleep disturbances related to menopause.

[1] The compound was initially developed by GSK, then NeRRe Therapeutics Ltd. on behalf of KaNDy Therapeutics Ltd. On 08 SEP 2020 Bayer AG acquired KaNDy Therapeutics Ltd. and continues development of elinzanetant.

The SAP describes the final analysis of the study. No statistical interim analysis will be performed. Table, figure and listing specifications are contained in a separate document.

Changes to the protocol-planned analyses are described in Section 4.8.

1.1 Objectives, Endpoints, and Estimands

Objectives and endpoints are listed in Table 1-1.

Table 1-1: Objectives and Endpoints

Objectives	Endpoints
Primary	
• To explore the efficacy of elinzanetant on sleep disturbances associated with menopause as determined by PSG	**Primary endpoint** • Change from baseline in WASO at week 4 as measured by PSG **Secondary endpoints** • Change from baseline in WASO at Week 12 as measured by PSG • Change from baseline in SE at week 4 as measured by PSG • Change from baseline in SE at week 12 as measured by PSG **Exploratory endpoints** • Change from baseline at week 4 and week 12 in PSG measures of sleep onset, duration, and continuity: LPS, TST, SOL, SOLN1, WTDS, EMWT, TWT, NAW, mean duration of awakenings, WASO each quarter of the night • Change from baseline at week 4 and week 12 in PSG measures of different sleep stage durations: time and percent in stage N1 sleep, time and percent in stage N2 sleep, time and percent in stage N3 sleep, time and percent in non-REM sleep, time and percent in REM sleep, REM onset latency from sleep onset • Change from baseline at week 4 and week 12 in other PSG measures: arousal index and number of stage shifts
Secondary	
• To explore the efficacy of elinzanetant on sleep disturbance associated with menopause as determined by patient reported outcomes	**Secondary endpoints** • Change from baseline in PROMIS SD SF 8b total score at week 4 • Change from baseline in PROMIS SD SF 8b total score at week 12 • Change from baseline in ISI total score at week 4 • Change from baseline in ISI total score at week 12 **Exploratory endpoints** • Change from baseline in subject-reported measures (as determined by the sleep diary (SD)) of sleep onset, duration, and continuity over time: sWASO, sLPS, sTST, sTIB, sSE, sNAW • Change from baseline in proportion of days with participants rating “quite a bit” or “very much” sleep disturbance due to HF as measured by HFDD over time

• To evaluate the safety of elinzanetant for the treatment of sleep disturbances associated with menopause	• TEAEs • Abnormal laboratory parameters
Exploratory	
• To explore the relationship between HFs and sleep disturbance as determined by PSG	• Frequency of HFs as measured by HFDD over time • WASO as measured by PSG over time
• To explore the efficacy of elinzanetant on the frequency of anxiety symptoms	• Change from baseline in GAD-7 total score over time
• To evaluate the efficacy of elinzanetant on the PGI-S and PGI-C	• Actual values of the PGI-C individual item scores over time • Change from baseline in PGI-S individual item scores over time
Other pre-specified	
• To evaluate variability in exposure in relation to the efficacy and safety for elinzanetant	• Systemic exposure of elinzanetant in plasma via sparse PK sampling
• To further investigate the study intervention and similar compounds (e.g., mode-of-action-related effects, safety) and to further investigate pathomechanisms deemed relevant to sleep disturbances associated with menopause.	• Various biomarkers (e.g., diagnostic, safety, pharmacodynamic, monitoring, or potentially predictive biomarkers)
• To investigate the efficacy of elinzanetant on sleep disturbance associated with menopause determined by Sleepiz One+	• Parameters captured remotely and continuously at home by Sleepiz One+ including but not limited to e.g., WASO, TST, SE

Estimands

The primary clinical question of interest for the primary objective is:

Does elinzanetant improve sleep disturbances associated with menopause? Specifically, what is the difference in mean change from baseline in WASO as measured by PSG at week 4 comparing elinzanetant to placebo in women with sleep disturbances associated with menopause assuming no study treatment interruption and discontinuation and no use of prohibited concomitant medications impacting sleep?

The attributes of the main estimand for the primary endpoint are as follows:

- Population: Women aged 40–65 with sleep disturbances associated with the menopause, further defined by the key inclusion/exclusion criteria.
- Variable: Change from baseline in WASO as measured by PSG at week 4.
- Treatment: Elinzanetant 120 mg or placebo.
- ICEs and strategies: Important ICEs to consider are listed in Table 1-2. (See Section 6.4 for details regarding identification of ICEs).

Table 1-2: Primary Estimand: ICEs and Strategies to Address Them

ICEs	Proposed Strategy	Data Handling Method	Interpretation
Temporary study treatment interruption[a]	Hypothetical	Discard data collected during the ICE	The treatment effect that would be if participants did not have any temporary treatment interruptions
Premature discontinuation of study treatment	Hypothetical	Discard data collected after the ICE	The treatment effect that would be if participants did not discontinue the randomized treatment
Intake of prohibited concomitant medication having impact on efficacy	Hypothetical	Discard data collected during the ICE	The treatment effect that would be if participants did not take prohibited concomitant medications impacting sleep

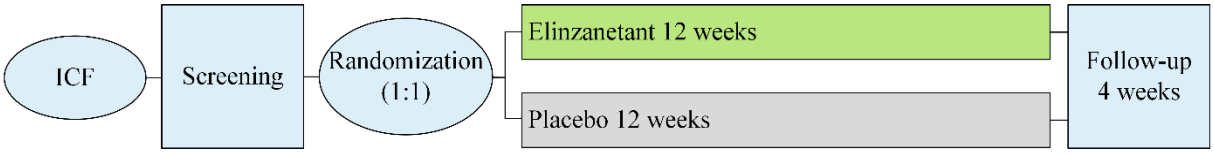
[a] Definition of temporary study treatment interruption:
Week 4 = Treatment taken <80% during day 1 to the last PSG night at week 4 (inclusive) OR treatment taken on <5 days during the last 7 days up to the last PSG night at week 4 (inclusive).

- Population level summary: Mean change from baseline in WASO as measured by PSG at week 4.

1.2 Study Design

This is a multi-center, multi-country, double-blind, randomized, parallel-group, placebo-controlled, pilot intervention study in women with sleep disturbances associated with the menopause.

Figure 1-1: Study Schema



ICF = signing of informed consent form.

The study duration will be about 23 weeks (plus potential washout period), including:

Screening: After giving informed consent, participants will be withdrawn from prohibited concomitant medications at SCR-1 visit if needed. Screening period consists of 3 visits.

Randomization: At the end of the screening period, the participants will be randomized into elinzanetant and placebo arms in a 1:1 ratio. Randomization will be stratified by the average weekly frequency of moderate to severe HFs (including nighttime HFs) over the 14 days prior to the respective eligibility check (<35 moderate to severe HFs per week, ≥35–<50 moderate to severe HFs per week, ≥50 moderate to severe HFs per week).

Intervention period: Intervention period will last about 12 weeks:

- 120 mg (2x 60 mg soft gel capsules) of elinzanetant or matching placebo orally once daily for 12 weeks

Follow-up: The intervention period will be followed by a 4-week safety follow-up period.

The visit frequency will be once every 1–2 weeks during the screening period (depending on the scheduling of PSG) and once every 4–5 weeks during the intervention and follow-up periods.

The study includes 3 stationary visits in a sleep laboratory. Each of these visits encompasses 2 consecutive nights. This results in a total of 6 nights in the sleep laboratory.

Primary analysis will be performed after database release after LPLV.

2. Statistical Hypothesis

The primary objective for the study is to estimate the treatment effect of elinzanetant compared to placebo in the change from baseline in WASO at week 4 as measured by PSG. No hypotheses are planned to be tested.

2.1 Multiplicity Adjustment

Not applicable.

3. Analysis Sets

For the purposes of analysis, the following analysis sets are defined in Table 3-1:

Table 3-1: Definition of the Analysis Sets

Participant Analysis Set	Description
Enrolled	All participants who signed the informed consent form.
Full Analysis Set (FAS)	All participants randomly assigned to study treatment.
Safety Analysis Set (SAF)	All participants randomly assigned to study treatment who were exposed to study treatment at least once.
Sleepiz One+ Analysis Set (SLAS)	All participants who provided consent for use of medical device and randomly assigned to study treatment with at least 1 non-missing Sleepiz One+ evaluation.

The FAS will be used to analyze endpoints related to the efficacy objectives and participants will be analyzed according to their randomized study treatment.

The SAF will be used to analyze the endpoints and assessments related to safety and participants will be analyzed according to the study treatment received.

The SLAS will be used to analyze the sleep endpoints assessed via the Sleepiz One+ device and participants will be analyzed according to their randomized study treatment.

Final decisions regarding the assignment of participants to analysis sets will be made during the review of study data and documented in the final list of important deviations, validity findings and assignment to analysis set(s).

4. Statistical Analyses

4.1 General Considerations

The statistical evaluation will be performed by using the software SAS (release 9.4 or higher; SAS Institute Inc., Cary, NC, USA) and/or R (version 3.5.2 or higher; R Foundation for Statistical Computing, Vienna, Austria).

All variables will be summarized using descriptive statistics as appropriate by study treatment. Continuous variables will be summarized using at least the following descriptive statistics: number of non-missing observations and missing data, arithmetic mean, std dev, median, minimum, and maximum. The geometric mean, geometric std dev and CV will be provided instead of the arithmetic mean and std dev for variables where lognormal distributions are assumed. Categorical data will be summarized with frequency tables with the number of observations and percentages. Confidence intervals (CI) will be 2 sided with a confidence level of 95% for descriptive statistics and 90% level from the MMRM. Treatment differences will be provided for elinzanetant vs placebo.

Variables recorded in the electronic case report form (eCRF) and relevant derived variables will be shown in data listings, whereby only randomized participants will be included. Data from screening failures will only be shown in the 'Screening failure' listing.

Number of decimal places for summary statistics will be the following:

Table 4-1: Number of Decimal Places for Summary Statistics

Statistic	Number of decimal places
Minimum, maximum	Same as original data
Mean, median, std dev	1 more than in original data
Frequencies (%)	1 digit

4.1.1 Handling of Missing Data

This section describes the general handling of missing data. This includes data that would be meaningful for the analysis in terms of the study objectives but were not collected. The rules for handling the data that do not exist or are not considered meaningful for the analysis because of an intercurrent event, are described in Section 4.2.2.1. No imputation of missing assessments will be performed for the secondary and exploratory endpoints unless it is specified otherwise.

All data will be presented in the participant data listing as they are recorded on the Case Report Form (CRF), i.e., partially missing data will appear as such.

4.1.1.1 Polysomnography (PSG)

Several parameters of sleep quality will be measured by on-site PSG assessments performed at screening, week 4 and week 12 with each visit consisting of 2 consecutive nights. See Section 6.2.1 for the list and definition of sleep parameters as measured by PSG.

For the analysis of PSG parameters, the mean value of the 2 consecutive nights at each visit will be calculated. If a participant only had one PSG night for a given visit, the data from the one night will be used for the analysis.

Further details regarding imputation rules for completely missing visits (i.e., no evaluable data from any nights) for WASO are described in Section 4.2.2.1. No imputation will be performed for other PSG parameters in case the visit is missing completely.

4.1.1.2 Sleepiz One+

Several parameters of sleep quality will be measured continuously at home and during the nights of the PSG assessments by the Sleepiz One + system from SCR-2 visit through to week 12/EoT visit. See Section 6.2.2 for the list and definition of sleep parameters as measured by Sleepiz One+.

No imputation of missing assessments will be performed for parameters measured by the Sleepiz One +.

4.1.1.3 Electronic Patient Reported Outcomes (ePROs)

There are two types of missing data for ePRO measures: missing data at the 'form' level and missing data at the 'item' level.

Form level missing refers to a participant missing an entire PRO assessment for a given time point (e.g., a given day (i.e., morning and evening) for the HFDD and a given week for the PROMIS SD SF 8b). In general, form level data may be missing due to participant's early withdrawal from the study, inability to evaluate an endpoint at a particular time point, or non-compliance.

By design of the eDiary, there will be no item level missing data for the respective ePRO questionnaires because the participants have to select an answer for an item in order to move on to the next item on the eDiary/electronic handheld device, with the exception for the eC-SSRS where the form can be submitted even if incomplete.

No imputation of missing assessments will be performed for any ePRO measures.

4.1.1.4 Adverse Events (AE) and Concomitant Medications (CM)

For computation of durations, i.e., time between start and end dates of certain events and concomitant medication intake, a complete date is necessary.

The following rule will be applied to impute partial missing start or end date of AE/ CM:

An imputation range will be defined as the earliest possible date (lower range) and the latest possible date (upper range) of the occurrence of AE/CM. If AE/CM end date is available, this will be used as the latest possible AE/CM occurrence date in the imputation range.

If the start of study treatment date falls within imputation range, incomplete AE/CM start date will be imputed as the start of study treatment date. Otherwise, the partially missing AE/CM start date will be imputed to the earliest date of the imputation range. Completely missing start date will not be imputed and will be considered treatment-emergent for AEs and concomitant for CMs.

Partially missing AE/CM end date will be imputed as the latest possible date of AE/CM imputation range. Completely missing end date will not be imputed and presumed to be ongoing.

4.1.1.5 Other Missing Data

The date of permanent discontinuation of study medication will be collected in eCRF. In case of missing date, it will be determined based on last medication intake date before the discontinuation using available eDiary data.

4.1.2 Data Rules

Baseline: Unless otherwise specified, baseline is defined as the latest available valid measurement (including unscheduled measurements) on or before the start date of study treatment (for treated subjects) / randomization date (for randomized but not treated subjects). If the last observation available prior to randomization is the measurement from the screening visit, this would be used as the baseline value.

Change from baseline: To calculate the absolute change from baseline, the baseline value will be subtracted from the value under treatment/follow-up, i.e.,

Absolute change = post baseline value – baseline value.

Some parameters will be additionally analyzed as percent change defined as:

Percent change = [(post baseline value – baseline value) / baseline value] x 100.

Visit confirmation date for HF eligibility check: Refers to the date, as captured on the electronic handheld device, when the eligibility for inclusion criteria of ≥ 20 average weekly frequency of moderate to severe HFs was assessed.

Laboratory values <X or >Y: For laboratory values lower than a limit of detection X (reported as <X), half the value of X will be used for analysis (i.e., for values reported as "<X", "<X.X"," <X.XX" or etc. the value for analysis will be derived by "X/2", "X.X/2","X.XX/2", etc.).

Difference between 2 values <X will be assigned a value of 0. Ratio between 2 values <X will be assigned a value of 1.

For values which are higher than a limit of detection Y (reported as, >Y), the value of Y will be used for analysis.

Repeated measurements at the same date: For PROMIS SD SF 8b, ISI, GAD-7, PGI-S and PGI-C, if more than one entry on the same date for any pre- and post-baseline visit is available, then the first entry will be used and all observations will be presented in the data listings.

Start of treatment: Refers to the first drug intake and will be identified based on the 'Study Intervention' eCRF entry.

Stratification factor: The stratification factor average weekly frequency of moderate to severe HFs (including nighttime HFs) (see Section 1.2) will be assigned in the IRT system during randomization. The average weekly frequency of moderate to severe HFs is based on HFDD data and will be calculated using the 14 days prior to the visit confirmation date for HF eligibility check. A minimum of 11 days should be available for the calculation of the average weekly value. In case participants are incorrectly stratified, the actual stratification will be used in the analysis. The actual average weekly frequency of moderate to severe HFs will be the value calculated in the eDiary/electronic handheld device. If more than 10% of

participants are incorrectly stratified, a sensitivity analysis may be performed using the assigned strata.

Study day 1: Defined as the day of start of treatment, or randomization day if participant did not take study treatment.

Study day: Study day will be calculated using study day 1 as the reference starting day. For observations on or after study day 1, the below formulae will be used:

Study day = date of observation – date of study day 1 + 1

For observations before study day 1, the formulae below will be used:

Study day = date of observation – date of study day 1

Study week: Study week will be calculated for study day >0 using the below formulae:

Study week = CEIL (study day/7) - 1

CEIL is a function in SAS that round up the value to the nearest integer. The calculation would follow the schedule of activities in the protocol (see Section 1.3 in protocol) where day 29 and day 85 with a plus 7-day visit window are specified for week 4 and 12 respectively. This would therefore calculate study day 1 to 7 as study week 0.

Time to event: The start date of treatment for treated participants or the randomization date for randomized participants who did not take study treatment will be used as the start date for the calculation of time to event. Therefore, the time of event will be defined as "date of event – start of treatment date (for treated participants) / randomization date (for randomized but not treated participants) +1".

Unscheduled assessments: Extra assessments (e.g., laboratory data or vital signs) associated with non-protocol visits will be included in listings, but not in the summary tables, unless otherwise specified.

4.1.3 Intercurrent Events (ICEs)

Important ICEs for this study are detailed in Table 1-2 and will be reviewed prior to study unblinding. Further details regarding the identification of ICEs are provided in Section 6.4.

For each ICE, the number of participants with an ICE will be summarized by week and by study treatment and overall. A participant may have more than one ICE and/or more than one occurrence of the same ICE. All ICEs observed will be summarized.

Boxplots of the observed WASO value separately for participants with no ICEs and participants with ICEs will be produced by visit and study treatment.

4.2 Primary Endpoint Analysis

4.2.1 Definition of Endpoint(s)

The primary endpoint is the change from baseline in WASO at week 4 as measured by PSG. See Section 6.2.1 for more details on the definition of WASO as measured by PSG.

PSG assessments will be performed on 2 consecutive nights during screening, at week 4 and at week 12. The baseline WASO value will be the mean value of the 2 consecutive nights at screening. Similarly, the week 4 value will be the mean value of the 2 consecutive nights at week 4. If a participant only had one PSG night for a given visit, the data from the one night will be used for the analysis.

Scheduled PSG assessments will be used but in case no scheduled assessment is available, unscheduled assessments while the participants are on treatment will be used and mapped with the following visit windows:

- if the unscheduled PSG fall within study day 22 to 45, then the assessment will be mapped to week 4. Likewise, if the unscheduled PSG fall within study day 78 to 103, then the assessment will be mapped to week 12.

If the participant discontinues treatment prematurely, the same visit windows will be used to map the EoT PSG assessment to the corresponding visit. EoT PSG assessments not mapped to week 4 or week 12 will be summarized together in the descriptive statistics but will be excluded from the model-based analysis.

When deriving study day, if the start time of the PSG recording is on or before 11:00, then the date of recording – 1 will be used as the date of observation.

The change from baseline in WASO to week 4 will be calculated as below:

Change from baseline = WASO mean value at week 4 – WASO mean value at baseline

The secondary endpoint, change from baseline in WASO at week 12 as measured by PSG, will be calculated analogous to the primary endpoint.

4.2.2 Main Analytical Approach

The primary estimand to answer the clinical question of the study is described in Section 1.1.

Analyses will be performed using the FAS.

The actual and change from baseline in WASO will be summarized using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI. In addition, descriptive statistics for actual values and change from baseline in WASO will be summarized by study treatment and individual nights (i.e., baseline night 1, baseline night 2, week 4 night 1, week 4 night 2, week 12/EoT night 1, week 12/EoT night 2).

The descriptive statistics for the actual and change from baseline in WASO by study treatment and week will also be summarized using the SLAS.

The change from baseline in WASO at week 4 and 12 will be analyzed using a MMRM with baseline value, study treatment, average weekly frequency of moderate-to-severe HFs (actual stratification factor <35 moderate to severe HFs per week, ≥35–<50 moderate to severe HFs per week, ≥50 moderate to severe HFs per week) and week included as covariates in the

model as well as the interaction term study treatment-by-week. An unstructured covariance structure will be used. If convergence cannot be attained, alternative structures will be attempted.

The LSM and standard error for each study treatment and the LSM difference of elinzanetant compared to placebo with corresponding 90% CI will be estimated from the model at week 4 and 12. Line plots of the LSM and 90% CI will be provided by study treatment and visit.

Model description including SAS code are provided in Table 4-2. SAS estimates the standard LSM over a balanced population by assigning equal weights to each category of the categorical covariates specified in the CLASS statement. The OBSMARGIN option within the LSMEANS statement will be specified to adjust the LSM according to the baseline distribution of the class covariates in the data.

Table 4-2: MMRM Model Description

Dependent variable:	change from baseline in WASO at week 4 and week 12
Covariates (continuous variables):	baseline WASO
Factors (class variables):	treatment, week, average weekly frequency of HFs (stratification factor)
Interaction terms:	treatment*week
Covariance structure:	Unstructured (UN) If convergence cannot be attained with the unstructured correlation matrix, the following alternative structures will be attempted in the specified order: autoregressive(1) (AR(1)), and compound symmetry (CS). The first structure in this list with which the model converges will be used. If AR(1) structure is used, a random subject intercept will also be included in the model. If multiple imputation (MI) is performed, then this will need to be applied for each imputed dataset (see Section 4.2.2.1).
Equation	$Y_{ijkl} = \mu + \beta x_i + t_k + r_l + v_j + (tv)_{jk} + s_i + \varepsilon_{ijkl}$ where Y_{ijkl} is the change from baseline in WASO to week j for subject i (with treatment k and baseline average weekly frequency of HFs l), μ is the intercept, β is the baseline covariate effect (baseline coefficient), x_i is the baseline WASO for subject i, t_k is the fixed effect of treatment k *(k=Elinzanetant, placebo)*, r_l is the fixed effect of average weekly frequency of HFs l *(l=<35 moderate to severe HFs per week, ≥35–<50 moderate to severe HFs per week, ≥50 moderate to severe HFs per week)*, v_j is the fixed effect of week j *(j=4, 12)*, $(tv)_{jk}$ is the interaction effect of treatment k by week j, $s_i \sim$ Normal $(0, \sigma_s^2)$ is the random effect of subject i *(only if AR(1) covariance structure is used)*, $\varepsilon_{i\,jkl} \sim$ Normal $(0, \sigma^2)$ represents the residual variance component with corr $(\varepsilon_{ij}, \varepsilon_{ij'}) = \rho_{jj'}, j \neq j'$.

SAS code:	PROC MIXED data=DATA method=ML; CLASS subject treatment average_HF week; MODEL change = baseline treatment average_HF week treatment*week /ddfm=KR outp=resid s; REPEATED week / subject=subject type=un; RANDOM subject; /*only if AR(1) covariance structure is used*/ LSMEANS treatment*week /alpha=0.1 cl diff obsmargins; ESTIMATE 'Elinzanetant - Placebo at Week 4' treatment 1 -1 treatment*week 1 0 -1 0 /alpha=0.1 cl; ESTIMATE 'Elinzanetant - Placebo at Week 12' treatment 1 -1 treatment*week 0 1 0 -1 /alpha=0.1 cl; ODS OUTPUT TESTS3=TYPE3_eff LSMeans=LSMEAN ESTIMATES=EST; RUN;

The assumption of normality in the MMRM will be assessed visually by residual plots (Q-Q plot and plot of residuals against predicted values). If multiple imputation (MI) is applied (see below Section 4.2.2.1), this will be assessed based on the observed data before multiple imputation.

4.2.2.1 Implementation of Strategies for Intercurrent Event Handling and Data Point Selection

Prior to performing the MMRM model, missing data will be handled in alignment with the estimand strategy for ICEs (described in Section 1.1).

The hypothetical strategy will be used to handle all of the ICEs in the primary estimand. This strategy would estimate the full efficacy potential of elinzanetant in the ideal situation if participants adhered to randomized study treatment up to week 4 without taking prohibited concomitant medications impacting sleep.

Any data that occur after premature discontinuation of study treatment will be discarded. For temporary treatment interruption and intake of prohibited concomitant medication, only data during the timeframe of the ICE will be discarded. For example, if a participant had a temporary treatment interruption at week 4 but not at week 12, only data at week 4 will be discarded while week 12 data will be utilized in the analyses. For prohibited concomitant medications, data collected during the concomitant medication period and within the washout period for that medication will be discarded. Since PSG assessments will be performed on 2 consecutive nights at each visit, only nights that are performed after premature discontinuation of study treatment or fall within the ICE timeframe will be discarded (e.g., if study treatment was taken on night 1 but not on night 2 then night 1 will be included and night 2 discarded). Details regarding how the ICEs are defined and will be identified are described in Section 6.4.

Given that all unobservable data (i.e., data discarded or missing during an ICE or missing values outside of the time frame of any ICE) will be modeled under the hypothetical scenario "if participant complied with treatment" (i.e., having no premature treatment discontinuation, no treatment interruption in the respective week and no intake of prohibited concomitant medications), it can be assumed that the missing or discarded data have a similar distribution as the observed data of participants who comply with treatment, given the covariates included

in the model (missing at random (MAR) mechanism). Therefore, missing or discarded data will be handled implicitly through using the MMRM model as specified in Section 4.2.2.

The MMRM model will exclude participants who have no post-baseline assessment from the analysis. Hence, if more than 5% of participants in the analysis set (FAS) have no post-baseline data (i.e., excluded from the MMRM model), Monte Carlo Markov Chain (MCMC) MI will be used to impute missing or discarded values prior to performing the MMRM model using the SAS code below.

```
PROC MI DATA=data OUT=outmi seed=22423 nimpute=500;

   MCMC CHAIN=multiple IMPUTE=full INITIAL=em;

   VAR change_wk4 change_wk12 baseline treatment average_HF1 average_HF2;

RUN;
```

Where average_HF1 and average_HF2 are indicator variables representing the level <35 moderate to severe HFs per week and ≥35–<50 moderate to severe HFs per week respectively of the variable actual stratification factor.

Each imputed dataset will be analyzed using the MMRM model specified in Table 4-2 to obtain an estimate of the treatment effect. The results will be combined using Rubin's rule (Rubin 1987) to obtain an overall mean change from baseline and overall estimated treatment effect. A total of 500 MI steps will be used.

4.2.3 Supplementary Analyses

Two supplementary estimands for the primary endpoint will be estimated:

1. The attributes are the same as the primary estimand (Section 1.1) except for the handling of the ICEs. The treatment policy strategy will be applied to handling temporary treatment interruption and intake of prohibited medications having impact on efficacy, while the hypothetical strategy if the participant remained untreated will be applied to premature discontinuation of study treatment.
2. The second supplementary estimand will be estimated if the assumption of normality is violated. The attributes are the same as the primary estimand, including the handling of ICEs (See section 1.1), except the variable will be change from baseline in log WASO as measured by PSG at week 4 and the population level summary will be the geometric mean ratio in WASO measured by PSG at Week 4.

4.2.3.1 First Supplementary Analysis

The information about ICEs and strategies to address them for the first supplementary estimand are provided in Table 4-3.

Table 4-3: First Supplementary Estimand: ICEs and Strategies to Address Them

ICEs	Proposed Strategy	Data Handling Method	Interpretation
Temporary study treatment interruption[a]	Treatment policy	Utilize the available data regardless of the occurrence of ICE	The treatment effect includes the effect of treatment interruptions
Premature discontinuation of study treatment	Hypothetical	Discard data collected after the ICE	The treatment effect that would be if participants who discontinue the randomized treatment remain untreated on randomized treatment
Intake of prohibited concomitant medication having impact on efficacy	Treatment policy	Utilize the available data regardless of the occurrence of ICE	The treatment effect includes the effect of prohibited concomitant medications impacting sleep

[a] Definition of temporary treatment interruption:
Week 4 = Treatment taken <80% during day 1 to the last PSG night at week 4 (inclusive) OR treatment taken on <5 days during the last 7 days up to the last PSG night at week 4 (inclusive).

A treatment policy will be applied to handle the ICE temporary study treatment interruption and intake of prohibited concomitant medication having impact on efficacy. A treatment policy approach for these ICEs is appropriate as this likely reflects the behavior of the target population. According to this strategy, all collected data should be utilized in the analysis irrespective of occurrence of the ICEs.

A hypothetical strategy will be applied to handle the ICE premature discontinuation of study treatment to estimate the treatment effect that would be if participants who prematurely discontinue study treatment remain untreated (i.e., do not receive the experimental study treatment until the planned assessment timepoint). This approach reflects what would happen if participants remained in the study after premature discontinuation of study treatment.

The same MMRM model as the primary estimand (Section 4.2.2) will be used and missing data will be handled in alignment with the first supplementary estimand strategy for ICEs prior to performing the MRMM model.

Although all participants are expected to be followed after ICEs temporary study treatment interruption and intake of prohibited concomitant medication, some missing data may occur. Missing values that occur while the participants continue on their randomized treatment are assumed MAR. Such missing values could be intermittent or monotone and will be imputed using a MCMC MI method.

The washout method (Wang et al. 2003) will be used to impute the unobservable data under the hypothetical scenario for the ICE premature discontinuation of study treatment. A placebo-effect is expected to be seen with the placebo group while the elinzanetant treatment effect will consist of both the placebo-effect and the effect of elinzanetant. The washout method effectively ‘washes out’ any pre-ICE treatment effect of elinzanetant in participants with an ICE randomized to elinzanetant, while modeling the mean placebo-effect. Missing data in participants randomized to placebo will be assumed MAR.

For participants randomized to elinzanetant, the imputation model will have the dependent variable as the endpoint measurement at week 4 with the baseline measurement and the stratification factor average weekly frequency of moderate-to-severe HFs as independent variables. The imputation model will be estimated only based on placebo participants who

remain in the study through week 4 and have available data at week 4. Similarly, week 12 values missing after an ICE will be imputed using the same model specification but with endpoint measurement at week 12 as the dependent variable.

The imputation will be done using a regression model (PROC MI with MONOTONE REG) including covariates described above for each time point and using the MNAR statement to estimate the model from placebo participants only.

To impute missing data for participants randomized to placebo, using MAR assumptions, the imputation will be done using a regression model including measurements from all timepoints before the timepoint being imputed in addition to baseline measurement and the stratification factor (PROC MI with MONOTONE REG). All placebo participants will be included in the imputation model.

PROC MI will be run with the random seed setting SEED=22423 to produce 500 imputed datasets.

Similar analysis steps (i.e., imputation, modeling and combining results) as detailed in Section 4.2.2 will be applied.

4.2.3.2 Second Supplementary Analysis

The MMRM model assumes that the errors are normally distributed. When the assumption of normal error distribution is violated, the model standard errors may be over-estimated.

The assumption of normality in the main analysis will be assessed visually by residual plots (Q-Q plot and plot of residuals against predicted values). If the plots indicate lack of normality, the main analysis will be repeated but WASO will be log-transformed and the variable change from baseline in log WASO at week 4 and week 12 will be analyzed.

The WASO value at each night will be log-transformed before calculating the mean value at each visit (see Section 4.2.1). The change from baseline in log WASO at week 4 and week 12 will be calculated as follows:

log WASO mean value at week 4/12 – log WASO mean value at baseline

The same analysis steps in alignment with the primary estimand strategy for ICEs (Section 4.2.2) will be implemented. In the MMRM model, baseline WASO value will be the baseline log WASO value instead.

The overall treatment effects will be back transformed into the original scale to report the geometric mean ratio to baseline for each study treatment and the geometric mean ratio for elinzanetant / placebo and corresponding 90% CIs. Q-Q plot and plot of residuals against predicted values will be produced to check the assumption of normality on the analysis of log-transformed data.

4.3 Secondary Endpoints Analysis

4.3.1 Secondary PSG Endpoints

- Change from baseline in WASO at week 12 as measured by PSG
- Change from baseline in sleep efficiency (SE) at week 4 as measured by PSG
- Change from baseline in SE at week 12 as measured by PSG

The secondary PSG endpoints will be calculated analogous to the primary endpoint (Section 4.2.1). See Section 6.2.1 for more details on the definition for each parameter as measured by PSG.

4.3.1.1 Change from Baseline in WASO at Week 12

Change from baseline in WASO at week 12 will be analyzed in the same MMRM model as the primary endpoint for the main analysis and for the supplementary analyses (Section 4.2.2. and 4.2.3).

4.3.1.2 Change from Baseline in Sleep Efficiency (SE) at Week 4 and 12

For the analysis of change from baseline in SE at week 4 and week 12, the ICEs will be handled according to the treatment policy and no missing data imputation is planned. The actual values and change from baseline in SE will be summarized using descriptive statistics by study treatment and week using the FAS.

The change from baseline over time by study treatment will be shown using line plots for means and 95% CI. In addition, descriptive statistics for actual values and change from baseline in SE will be summarized by study treatment and individual nights (i.e., baseline night 1, baseline night 2, week 4 night 1, week 4 night 2, week 12 night 1, week 12 night 2).

The descriptive statistics for the actual and change from baseline in WASO by study treatment and week will also be summarized using the SLAS.

4.3.2 Secondary PRO Endpoints

- Change from baseline in PROMIS SD SF 8b total score at week 4
- Change from baseline in PROMIS SD SF 8b total score at week 12
- Change from baseline in ISI total score at week 4
- Change from baseline in ISI total score at week 12

For all secondary PRO endpoints, the ICEs will be handled according to the treatment policy. No missing data imputation is planned for these endpoints. Analyses will be performed using the FAS.

4.3.2.1 Patient-reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b (PROMIS SD SF 8b)

The PROMIS SD SF 8b includes 8 items assessing sleep disturbance over the past 7 days (Yu et al, 2011). Items assess sleep quality, sleep depth and restoration associated with sleep, perceived difficulties with getting to sleep or staying asleep and perceptions of the adequacy of and satisfaction with sleep. Participants respond to the items on a 5-point scale (scored from 1 to 5) from “not at all”, “never”, “very poor” to “very much”, “always” or “very good”. Four of the items are scored reversely.

Participants’ responses to the 8 items of the instrument are scored on a 1-5 numeric rating scale and will be aggregated to derive total raw scores ranging from 8-40 with higher scores indicating greater severity of sleep disturbance. All PROMIS SD SF 8b items must be answered to produce a valid total score.

These total raw scores will be converted into T-scores for comparison with population norms (United States general population). The conversion table is provided in Table 4-4. (PROMISSleep Disturbance – Scoring Manual).

Table 4-4: PROMIS Sleep Disturbance 8b - Conversion Table

Sleep Disturbance 8b *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.9	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE = Standard Error on T-score metric

PROMIS SD SF 8b is collected weekly per the SOA in the protocol (Section 1.3 of protocol). Weekly and EoT assessments will be assigned to study week taking the recall period of 7 days into account, i.e., an assessment will be assigned to that study week that covers the majority of the recall period (see Table 4-5). For example, an assessment performed on day 8 will be assigned to week 0, as the recall period covers days 1 to 7, which matches week 0. An assessment performed on day 12 will be assigned to week 1, as the recall period covers days 5 to 11, of which 3 days belong to week 0 and 4 days belong to week 1. If more than one assessment is assigned to a week, e.g., participant completes a weekly assessment and very shortly has a protocol scheduled visit (including EoT), the latest available assessment will be used in the analysis.

The baseline value will be the earliest assessment on day 1 or day 2. If both day 1 and day 2 assessment is not available, the baseline value will be the latest available assessment on or

before day 4. Only assessments where ≥ 4 days of the 7-day recall period are outside of the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy will be included. Details regarding the prohibited medications together with the impact duration are described in Section 6.3.

Table 4-5: PROMIS Study Week Mapping

PROMIS assessment day (inclusive)	Study Week
≤ 4	Baseline
5 - 11	Week 0
12 - 18	Week 1
19 - 25	Week 2
26 - 32	Week 3
33 - 39	Week 4
40 - 46	Week 5
47 - 53	Week 6
54 - 60	Week 7
61 - 67	Week 8
68 - 74	Week 9
75 - 81	Week 10
82 - 88	Week 11
89 - 95	Week 12

The absolute values and change from baseline values for PROMIS SD SF 8b total T-scores and total raw scores will be summarized using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI. In addition, the number and percentage of each response for each individual item will be summarized using frequency tables by study treatment and week. Stacked bar charts will also be produced showing the composition of responses for each item by treatment group and week.

4.3.2.2 Insomnia Severity Index (ISI)

The ISI is a 7-item instrument that quantifies the participants perception of insomnia severity in the last 2 weeks (Bastien et al, 2001) and is collected during screening, at week 4 and at week 12/EoT. The items refer to: severity of sleep onset, sleep maintenance and early morning wakening problems, satisfaction with sleep pattern, noticeability of sleep problems by others, distress caused by the sleep difficulties and interference of sleep difficulties with daytime functioning. It is scored on a 5-point Likert scale from 0 to 4. The scores for each

item are summed to produce the total score (maximum of 28) which allow categorization of severity of insomnia.

Total score severity categories:

- 0–7 = No clinically significant insomnia
- 8–14 = Subthreshold insomnia
- 15–21 = Clinical insomnia (moderate severity)
- 22–28 = Clinical insomnia (severe)

The baseline value will be the earliest assessment on day 1 or day 2. If both day 1 and day 2 assessment is not available, the baseline value will be the latest available assessment on or before day 4. Only assessments where ≥ 8 days of the 14-day recall period are outside of the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy will be included. Details regarding the prohibited medications together with the impact duration are described in Section 6.3.

The absolute values and change from baseline values for ISI total score will be summarized using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI.

The ISI severity categories will be summarized using frequency tables by study treatment and week. Stacked bar charts showing the composition of severity categories at each week by treatment group will be produced.

In addition, the number and percentage of each response for each individual item will be summarized using frequency tables by study treatment and week. Stacked bar charts will also be produced showing the composition of responses for each item by treatment group and week.

4.4 Exploratory/Other Endpoints Analysis

For all exploratory endpoints, the ICEs will be handled according to the treatment policy. No missing data imputation is planned for these endpoints. Analyses will be performed using the FAS.

4.4.1 Other PSG Exploratory Endpoints

- Change from baseline at week 4 and week 12 in PSG measures of sleep onset, duration, and continuity: LPS, TST, SOL, SOLN1, WTDS, EMWT, TWT, percentage of TWT, NAW, mean duration of awakenings, WASO by each quarter of the night.
- Change from baseline at week 4 and week 12 in measures of different sleep stage duration: time and percent in stage N1 sleep, time and percent in stage N2 sleep, time and percent in stage N3 sleep, time and percent in non-REM sleep, time and percent in REM sleep, REM onset latency from sleep onset.
- Change from baseline at week 4 and week 12 in other PSG measures: arousal index and number of stage shifts.

The calculation of the change from baseline to week 4 and 12 will be analogous to the primary endpoint (Section 4.2.1). See Section 6.2.1 for more details on the definition for each parameter as measured by PSG.

The actual values and change from baseline will be summarized using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI. In addition, descriptive statistics for actual and change from baseline will be summarized by study treatment and individual nights (i.e., baseline night 1, baseline night 2, week 4 night 1, week 4 night 2, week 12 night 1, week 12 night 2).

4.4.2 Sleep Diary (SD) Related Endpoints

- Change from baseline in subject-reported measures (as determined by SD) of sleep onset, duration, and continuity over time: sWASO, sLPS, sTST, sTIB, sSE, sNAW

Participant's assessment of sleep disturbance will be recorded electronically daily using the Sponsor developed SD. The SD is completed in the morning after waking up (calendar day) to assess the participant's sleep disturbance during the previous night (diary day). The SD consists of 6 items that assess the total sleep time duration by asking for time of going to bed, falling asleep and waking up, as well as duration spent in bed, number and duration of awakenings during the night.

SD assessments that do not logically make sense will be excluded from the analysis but will still be included in the listings. Assessments that satisfy one or more of the below criteria will be excluded:

1. If time taken to fall asleep ≥ duration between time going to bed and time waking up
2. If total duration of awakenings ≥ duration between time going to bed and time waking up
3. If total duration of awakenings ≥ duration between time going to bed and time waking up – time taken to fall asleep
4. If time going to bed is between 06:00 and 16:00 (inclusive), and duration between time going to bed and time waking up is <1 hour or ≥ 16 hours.
5. If time waking up is between 12:00 – 23:59 (inclusive), and duration between time going to bed and time waking up is <1 hour or ≥ 16 hours.
6. If time going to bed = time waking up
7. If time going to bed = time getting out of bed

Subject-reported parameters to be analyzed will be calculated using items from the SD as detailed in Table 4-6:

Table 4-6: Definition of Subject-Reported Parameters from SD

Subject-reported parameter	SD item(s)
sWASO (mins)	Item 4: total duration of awakenings
sLPS (mins)	Item 2: time taken to fall asleep
sTST (hours)	Time between going to bed (Item 1) and waking up in the morning (Item 5) minus the time taken to fall asleep (Item 2) and total duration of awakenings (Item 4)
sTIB (hours)	Time between going to bed (Item 1) and getting out of bed (Item 6*)

sSE (%)	(sTST / sTIB) * 100
sNAW	Item 3: total number of times woke up in the night

* If time getting out of bed (item 6) is before time waking up in the morning (item 5), then time getting out of bed will be set as the same time as waking up.

For the analysis, weekly mean values will be used. The mean daily value for each study week will be calculated by using available data during that week as:

For a particular parameter:

$$\frac{\text{total value of the parameter during that week}}{\text{total number of available days with morning data for that parameter during that week}}$$

Study week will be assigned using the definition in Section 4.1.2 using diary day as the date of observation. Data entered on the morning after the start date of treatment / randomization date will not be used in the calculation of baseline value but will be considered post-baseline.

The mean daily value can be derived if a value is provided for at least 4 out of 7 days with morning data, otherwise the value will be set as missing. The baseline value will be calculated similarly using all available days with morning data during the 14 days prior to starting treatment. A minimum of 7 days should be available for the derivation of the baseline value. Only diary days outside of the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy will be included in baseline derivation. Details regarding the prohibited medications together with the impact duration are described in Section 6.3.

The mean daily value and change from baseline in mean daily value for each parameter will be analyzed using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI.

In addition, the mean daily value from the PSG nights will be summarized descriptively by study treatment and week. The mean value will be calculated using the same rule to calculate the primary endpoint (Section 4.2.1), i.e., if there are sleep diary data for both PSG nights, the mean value of the 2 nights will be used; if only 1 night is available, data from that one night will be used.

4.4.3 Hot Flash Daily Diary (HFDD)

- Frequency of HFs as measured by HFDD over time
- Change from baseline in proportion of days with participants rating “quite a bit” or “very much” sleep disturbance due to HF as measured by HFDD over time

Participants’ assessments of HFs will be recorded electronically twice daily using the sponsor developed HFDD. The HFDD is completed in the morning after waking up (morning diary) and each evening at bedtime (evening diary) on the hand-held device. The HFDD assess the number of mild, moderate, and severe HFs experienced during the day (evening diary) and during the night (morning diary). Therefore, a diary day consists of the evening entry and the morning entry of the subsequent day.

Mild HFs are defined as a "sensation of heat without sweating", moderate HFs are defined as a "sensation of heat with sweating, but able to continue activity", and severe HFs are defined as a "sensation of heat with sweating, causing cessation (stopping) of activity".

The following daily HF assessments will be aggregated to a mean daily frequency for each study week:

- moderate to severe nighttime HFs (morning diary)
- mild, moderate, and severe nighttime HFs (morning diary)
- moderate to severe HFs (morning and evening diary)
- mild, moderate, and severe HFs (morning and evening diary)

The mean daily frequency of moderate to severe nighttime HFs for a study week will be calculated as

$$\frac{\text{total number of nighttime moderate to severe HFs during that week}}{\text{total number of available days with morning data during that week}}$$

The calculation of mild, moderate, and severe nighttime HFs, moderate to severe HFs and mild, moderate, and severe HFs will be calculated similarly to the above.

Study week will be assigned using the definition in Section 4.1.2. using the diary day as the date of observation. Data entered on the morning after the start date of treatment / randomization date will not be used in the calculation of baseline value but will be considered post-baseline for nighttime HFs. For the mean daily frequency using both the morning and evening diary (i.e., moderate to severe HFs and mild, moderate, and severe HFs), diary day 2 – 8 will be used for week 0.

A minimum of 4 days should be available within a study week for the derivation of mean daily frequency, otherwise it will be set to missing for that week. The baseline value will be calculated similarly using all available days with data during the 14 days prior to starting treatment. A minimum of 7 days should be available for the derivation of the baseline value. Only diary days outside of the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy will be included in baseline derivation. Details regarding the prohibited medications together with the impact duration are described in Section 6.3.

A day would be considered as available if morning entry (of the subsequent day) is available when only morning diary is used, or if either evening entry or morning entry (of the subsequent day) is available when both evening and morning diaries are used.

The relationship between the frequency of HFs and WASO as measured by PSG will be explored. The PSG value as defined in Section 4.2.1 will be used in the analysis. Scatterplots showing the below will be produced separately for the actual values, change from baseline and percent change from baseline:

- mean daily frequency of moderate to severe nighttime HFs against WASO as measured by PSG
- mean daily frequency of mild, moderate and severe nighttime HFs against WASO as measured by PSG

- mean daily frequency of moderate to severe HFs against WASO as measured by PSG
- mean daily frequency of mild, moderate and severe HFs against WASO as measured by PSG

The mean daily HFs for the same week the week 4 and week 12/EoT PSG assessments fall in will be plotted against each other. Scatterplots will be produced combining all visits on the same plot. In addition, the number of observations and Spearman's rank correlation coefficient and Pearson correlation coefficient with corresponding 95% CI will be computed and presented on the scatterplot.

The mean daily frequency of moderate to severe HFs (morning and evening diary) and mean daily frequency of mild, moderate, and severe HFs (morning and evening diary) as well as the change from baseline values will be analyzed using descriptive statistics by study treatment and week. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI.

Sleep disturbances due to HFs are also assessed every morning as part of the HFDD. Participants provide a rating on their severity of sleep disturbances due to HFs in the previous night as

- 'not at all'
- 'a little bit'
- 'somewhat'
- 'quite a bit'
- 'very much'

To calculate the proportion of days with "quite a bit" or "very much" sleep disturbance due to HFs, the number of days with participants having reported "quite a bit" or "very much" sleep disturbance will be used.

The proportion of days with quite a bit or worse sleep disturbances due to HFs for each study week will be calculated using available data during that particular week as

$$\frac{\text{number of days with rating 'quite a bit' or 'very much' sleep disturbance due to HFs during the week}}{\text{total number of available days with morning data during that week}}$$

Study week assignment and number of available days for calculation of baseline and weekly values will follow the same rules as above for mean frequency of moderate to severe nighttime HFs.

The proportion of days with participants having reported "quite a bit" or "very much" sleep disturbance due to HFs as well as the change from baseline will be summarized using descriptive statistics by study treatment and week. The change from baseline over time will be shown using line plots for means and 95% CI by treatment group.

The proportion of days with participants having reported "not at all" or "a little bit" sleep disturbance due to HFs as well as the change from baseline will also be summarized using descriptive statistics by study treatment and week. The change from baseline over time will be

shown using line plots for means and 95% CI by treatment group. The proportion of days with “not at all” or “a little bit” sleep disturbance due to HFs will be calculated similarly as proportion of days with “quite a bit’ or “very much” sleep disturbance.

4.4.4 General Anxiety Disorder-7

- Change from baseline in GAD-7 total score over time

The GAD-7 is a 7-item instrument that quantifies the symptom severity in generalized anxiety disorder in the last 2 weeks (Spitzer et al, 2006) and is collected during screening, at week 4 and at week 12/EoT. Concepts assessed are the frequency of being anxious, worried, having troubles relaxing, being restless, annoyed or irritated and feeling afraid. The items are scored on a 4-point Likert scale from 0 to 3 (0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly every day). The scores are summed to produce the total score (maximum of 21), with higher scores indicating a higher frequency of anxiety symptoms.

The total score can be categorized into the below level of anxiety severity:

- 0–4 = None/minimal anxiety
- 5–9 = Mild anxiety
- 10–14 = Moderate anxiety
- 15–21 = Severe anxiety

The item ‘If you checked of any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?’ from the PHQ-9 is also assessed as part of the GAD-7 form on the ePRO device. Participants respond using a 4-point response scale ‘Not difficult at all’ to ‘Extremely difficult’.

The absolute values and change from baseline values for GAD-7 total score will be summarized using descriptive statistics by study treatment and week. The baseline value will be the earliest assessment on day 1 or day 2. If both day 1 and day 2 assessment is not available, the baseline value will be the latest available assessment on or before day 4. Only assessments where ≥ 8 days of the 14-day recall period are outside of the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy will be included. Details regarding the prohibited medications together with the impact duration are described in Section 6.3. The change from baseline over time by study treatment will be shown using line plots for means and 95% CI.

The GAD-7 anxiety severity categories will be summarized using frequency tables by study treatment and week. Stacked bar charts showing the composition of anxiety levels at each week by treatment group will be produced.

In addition, the number and percentage of each response for each individual item, as well as the item from PHQ-9, will be summarized using frequency tables by study treatment and week. Stacked bar charts will also be produced showing the composition of responses for each item by treatment group and week.

4.4.5 Patient Global Impression of Severity (PGI-S) and Change (PGI-C)

- Change from baseline in PGI-S individual item scores over time
- Absolute values of the PGI-C individual item scores over time

The PGI-S includes 2 items assessing sleep problems over the past week using a 5-point response scale ("No sleep problems" to "Very severe" for the PGI-S item 1 (severity of sleep problems) and "None of the time" to "Nearly all of the time" for PGI-S item 2 (time spent awake after first falling asleep)). Each item is answered independently.

The PGI-C includes 2 items assessing change in sleep problems since the participant started taking the study medication. A 5-point response scale is used ranging from "much better" to "much worse" for PGI-C item 1 (change in severity of sleep problems) and "much less" to "much more" for PGI-C item 2 (change in time spent awake). Each item is answered independently.

Both instruments, the PGI-S and the PGI-C are self-invented by the sponsor strongly following regulatory guidance (FDA 2018) and are collected weekly as per the SOA in the protocol (Section 1.3 of protocol). Weekly and EoT assessments will be assigned to study week using the same mapping as PROMIS (see Section 4.3.2.1), i.e., accounting for the recall period of 7 days, and the impact duration of prohibited hormonal therapies and medication potentially confounding efficacy for baseline derivation. If more than one assessment is assigned to a week, e.g., participant completes a weekly assessment and very shortly has a protocol scheduled visit (including EoT), the latest available assessment will be used in the analysis.

The answers to the 2 individual items for PGI-C and PGI-S will be summarized in frequency tables by study treatment and week. The change from baseline in PGI-S individual scores will also be analyzed using shift tables (observed frequencies at baseline versus post-baseline weeks).

4.5 Safety Analyses

The analyses described in this section will be presented for the SAF.

4.5.1 Extent of Exposure

Treatment duration will be defined as the number of days from the day of first study intervention intake up to and including the day of last study intervention intake and will be summarized using descriptive statistics by study treatment. Treatment duration will be presented based on the data collected via the eCRF. If the date for end of exposure is not available in eCRF, the last day of intervention intake from eDiary will be used to determine the end of exposure.

The compliance (as percentage) will be calculated as:

100 * Number of capsules taken / Number of planned capsules

The number of planned tablets is calculated as:

Treatment duration * 2

All capsules, including the placebo capsules, will be counted. For participants who withdraw prematurely from the study treatment, compliance will be calculated up to the time of last dose. The compliance will be summarized descriptively by study treatment. In addition, percent of compliance will be categorized into 3 groups, <80%, 80 to 120% >120%, and the categories will be summarized by study treatment.

Compliance will be calculated and presented twice, once based on the ePRO daily instrument 'Study medication intake documentation' and once based on the eCRF 'Drug Accountability' and 'Study Intervention' pages. If the 'Study medication intake documentation' on the second PSG night at week 12/EoT is missing but the last treatment intake on the eCRF 'Study intervention' pages indicate that the participant took treatment on the night, the 'Number of study medication capsules taken' on the 'Study medication intake documentation' will be imputed with 1 capsule.

The summaries for treatment duration and compliance will be presented for the overall intervention period, (i.e., 12 weeks).

4.5.2 Adverse Events (AEs)

All AEs will be summarized using MedDRA (the current version at the time of analysis) PTs grouped by SOC.

Any AEs related to study procedures recorded after signing of informed consent but prior to randomization will be considered as pre-treatment AEs. AEs that occurred or worsened after the first dose of study intervention up to 14 days after the last dose of study intervention will be considered as treatment-emergent AEs (TEAEs). AEs that occurred 15 days after last dose will be considered as post-treatment AEs.

Partially missing onset AE date will be imputed following a worst-case approach as described in Section 4.1.1.4.

If the severity of a previously reported event worsens or if the relationship to study intervention/protocol required procedures changes, two separate events should be reported.

Worsening of an AE is defined as follows:

- AE intensity is worsened (e.g., moderate to severe)
- AE changed to a serious event
- AE ends with death

In case an AE starts before the date of last study intervention intake + 14 days and worsens after last study intervention intake +14 days, it will be considered as two AEs, a TEAE and a post-treatment AE.

In case of events with different intensity within a participant, the maximum reported intensity will be used. If the same event is reported as both unrelated and related to the study intervention within a participant, the event will be considered as related to study intervention. If the study intervention relationship is missing, the event will be considered as being related to the study intervention. For AEs, relationship to study intervention refers to elinzanetant and placebo only.

Overall summaries of the number of participants with pre-treatment AEs, TEAEs, and post-treatment AEs will be generated by study treatment.

The number of participants with the following AEs will be summarized by study treatment using PTs grouped by SOC:

- pre-treatment AEs
- post-treatment AEs

- SAEs (includes both treatment-emergent and post-treatment)
- TEAEs, TEAEs of special interest, TEAEs resulting in discontinuation of study intervention, TEAEs by maximum intensity
- treatment-emergent study intervention related AEs, treatment-emergent study intervention-related AEs by maximum intensity, treatment-emergent study intervention-related AEs resulting in discontinuation of study intervention
- treatment-emergent SAEs, treatment-emergent SAEs of special interest, treatment-emergent SAEs resulting in discontinuation of study intervention, treatment-emergent SAEs by maximum intensity, treatment-emergent study intervention-related SAEs, treatment-emergent study intervention-related SAEs by maximum intensity
- AEs with fatal outcome (includes both treatment-emergent and post-treatment), TEAEs with fatal outcome

Listings of treatment-emergent SAEs, and AEs with fatal outcome will be provided.

Adverse events of special interest (AESI)

The following are defined as AESI in this study, if the event takes place after the first intake of study intervention:

1. Any condition triggering close liver observation (as listed in Protocol Section 8.4.6): identified by Standardized MedDRA Query (SMQ) searches and AESI 'tick-box' via eCRF
2. Post-menopausal bleeding: identified by MedDRA Labeling Grouping (MLG) and PT searches

See details in Section 6.5 for SMQ, PT and MLG search criteria.

4.5.3 Additional Safety Assessments

For laboratory, liver monitoring, and vital signs the number of participants whose baseline value is a measurement while taking or washing-out prohibited hormonal therapies or medications potentially confounding efficacy during the screening period will be reported in a table.

4.5.3.1 Laboratory Data

The number of participants with treatment-emergent (i.e., 14 days from the last treatment intake) high or low abnormal laboratory values will be summarized for each laboratory parameter by study treatment. Both scheduled and unscheduled laboratory measurements will be considered.

Continuous laboratory parameter values including the change from baseline will be summarized by visit and study treatment. Frequency tables for categorical laboratory parameter will be summarized by visit and study treatment.

Central laboratory will be considered for descriptive analysis. Local laboratory measurements will be listed, if available.

4.5.3.2 Liver Monitoring

The following parameters will be investigated in addition to the standard lab presentations:

- Aspartate aminotransferase (AST) (in U/L),
- Alanine aminotransferase (ALT) (in U/L),
- Alkaline phosphatase (ALP) (in U/L),
- Total bilirubin in serum (TB) (in mg/dL)
- International normalized ratio (INR).

Frequency tables, presenting number and percentage of participants, by treatment group for baseline and post-baseline (scheduled and unscheduled measurements) will be presented for the following categorizations (where ULN stands for "Upper Limit of Normal"):

- For ALT and AST, separately:
  - ≥1×ULN, ≥3×ULN, ≥ 5×ULN, ≥ 8×ULN, ≥ 10×ULN, ≥ 20×ULN
- For ALT and AST combined (if at least one of ALT and AST falls into the category):
  - ≥ 3×ULN, ≥ 5×ULN, ≥8×ULN, ≥ 10×ULN, ≥ 20×ULN
- For TB:
  - ≥1×ULN, ≥2×ULN, ≥ 5×ULN, ≥8×ULN
- For ALP:
  - ≥1.5×ULN, ≥2.0×ULN, ≥3.0×ULN
- For INR:
  - ≥1.5, ≥2

Frequency tables presenting number and percentage of participants by treatment group for post-baseline (scheduled and unscheduled measurements) will be presented for the following combinations of ALT/AST and TB or INR,

Relative to ULN:

- ALT or AST ≥3×ULN and ≥1.5×ULN in TB
- ALT or AST ≥ 3×ULN and ≥2×ULN in TB
- ALT or AST ≥3×ULN followed by ≥2×ULN in TB (measured within 30 days afterwards) (Hy's Law criteria).
- (ALT or AST ≥3×ULN) and ≥1.5 for INR
- ALT or AST ≥5×ULN for more than 2 weeks
- ALT or AST ≥8×ULN
- ALT or AST ≥ 3×ULN with the appearance of any signs or symptoms on liver event - clinical signs and symptoms eCRF page
- ALP ≥ 2×ULN and TB ≥2×ULN

Relative to baseline (BL):

- (ALT ≥3×BL or AST ≥3×BL) and TB ≥2×BL

- ALT ≥3×BL or AST ≥3×BL
- ALT ≥5×BL or AST ≥5×BL
- (ALT ≥2×BL or AST ≥2×BL) and TB ≥2×BL
- ALP ≥2×BL and TB ≥2×BL

Time to event analysis

Cumulative incidence estimates for the time to first occurrence of ALT ≥ 3×ULN and first occurrence of ALP ≥ 3×ULN will be derived. If no such an increase is observed, the observation is censored at the last visit date. Tables with the number of participants under risk, cumulative number of participants with ALT ≥3×ULN and ALP ≥ 3×ULN, and estimated probability for an event including 95% CIs (two-sided) will be presented. Furthermore, cumulative incidence curves will be provided by treatment group. If a participant does not have any post-baseline data, she will be censored at baseline.

Figures

Hepatocellular DILI screening plot will be displayed to identify cases of possible serious hepatocellular DILI. In the plot each patient is plotted based on their maximum postbaseline TB (y-axis) and transaminase values (ALT or AST, whichever is higher). Together with the plot, a frequency table for patients in each quadrant (Potential Hy's Law, Cholestasis, Temple's corollary) will be shown.

Cholestatic drug-induced liver injury screening plot will be shown to identify significant ALP elevation in the setting of hepatic dysfunction. Maximum postbaseline TB is plotted against maximum postbaseline ALP. Similar to above, a frequency table for patients in each quadrant (TB ≥ 2xULN and ALP ≥ 2xULN, TB ≥ 2xULN and ALP < 2xULN, TB < 2xULN and ALP ≥ 2xULN) will be shown.

Individual patient presentations

If a participant meets the criteria for close liver observation (as specified in Section 10.5.1 of the protocol) at any time point, a plot for her individual time course in the following laboratory parameters will be presented: 1) ALT, AST, TB, and ALP relative to ULN over time and 2) absolute values for INR over time. It will be indicated within the plot on which days the study drug was taken (i.e. start and stop dates of treatment intake), in addition all AEs and CM 6 months (182 days) prior to the first onset of close liver observation will be shown in the plot 1. Furthermore, listings will be provided for INR (absolute values) and for liver-related parameters, i.e. ALT, AST, TB, ALP, GGT, direct bilirubin, CK and LDH (with results relative to ULN).

4.5.3.3 Vital Signs

Vital sign (pulse rate, systolic blood pressure and diastolic blood pressure) will be taken at screening, week 4 and week 12/EoT, and measurements maybe repeated in case of abnormal results. If more than 1 measurement are available on the same date, the average will be calculated and summarized descriptively including the change from baseline by visit and study treatment.

4.5.3.4 Electronic Columbia-Suicide Severity Rating Scale (eC-SSRS)

Suicidal ideation and behavior will be monitored by eC-SSRS questionnaire.

The eC-SSRS assesses lifetime suicidality during an initial baseline evaluation, and then prospectively monitors ideations and behaviors at subsequent follow-up assessments. The eC-SSRS assesses the severity and frequency of suicidal ideation and behavior.

eC-SSRS outcomes will be summarized by study treatment for each visit (baseline, week 4 and 12) using descriptive statistics.

4.5.3.5 Pregnancy Test

Pregnancy test results will be listed.

4.6 Other Analyses

Other pre-specified objectives in this study are:

- To evaluate variability in exposure in relation to the efficacy and safety for elinzanetant
- To further investigate the study intervention and similar compounds (e.g., mode-of-action-related effects, safety) and to further investigate pathomechanisms deemed relevant to sleep disturbances associated with menopause.
- To investigate the efficacy of elinzanetant on sleep disturbance associated with menopause determined by Sleepiz One+

These will be evaluated accordingly by:

- Systemic exposure of elinzanetant in plasma via sparse PK sampling
- Various biomarkers (e.g., diagnostic, safety, pharmacodynamic, monitoring, or potentially predictive biomarkers)
- Parameters captured remotely and continuously at home by Sleepiz One+ including but not limited to e.g., WASO, TST, SE

4.6.1 Pharmacokinetics

PK analyses will be described in a separate document and the results will be presented outside of the CSR. Only concentration data will be provided in a listing.

4.6.2 Biomarkers

The analyses of various biomarkers for diagnostic, safety, pharmacodynamic, etc. will be described in a separate document and the results will be presented outside of the CSR.

4.6.3 Sleepiz One+

The analyses of sleep parameters derived from Sleepiz One + monitor will be presented for the SLAS.

The Sleepiz One+ monitor will be provided to participants at screening to continuously collect sleep data at home in addition to the nights of PSG assessment. Parameters derived from Sleepiz One+ to be analyzed will be:

- WASO (mins)
- SE (%)
- TST (hours)

- SOL (mins)
- Wake up latency (mins)
- Awakening count
- Average awakening (per hour of sleep)
- Out of bed count
- Time in bed (hours)
- Respiratory irregularity index (RII) (per hour of sleep)
- Heart rate per minute (min, max, mean, median, 10^{th} and 90^{th} percentile)
- Breathing rate per minute (min, max, mean, median, 10^{th} and 90^{th} percentile)

See Section 6.2.2 for more details on the definition of each parameter.

The parameters listed above are provided by the vendor as nightly values. For the analysis, the weekly mean daily value will be used. The mean value for each study week will be calculated using available data during that week as:

For a particular parameter:

$$\frac{\text{total value of the parameter during that week}}{\text{total number of available days with data for that parameter during that week}}$$

Study week will be assigned using the definition in Section 4.1.2 using the start date of the recording as the date of observation. If the start time of the recording is on or before 11:00, then the date of recording – 1 will be used as the date of observation. The mean daily value can be derived if a value is provided for at least 4 out of 7 nights, otherwise the mean daily value will be set as missing. The baseline value will be calculated similarly using all available days during the 14 days prior to starting treatment. A minimum of 7 days should be available for the derivation of the baseline value.

In single cases of concern that data was not transmitted properly during the participant's Sleepiz One+ use at home, the device will be switched on at site for 24 hours (after the participant returns the device) to transmit potentially stored data. Data recorded during this time will not be analyzed or listed. Therefore, data on or after last PSG assessment date + 1 or last treatment date + 1 or EoT visit date (whichever is the latest) will be excluded from the analysis.

The mean daily value and change from baseline in mean daily value for each parameter will be analyzed using descriptive statistics by study treatment and week. The mean daily value and change from baseline over time by study treatment will be shown using line plots.

In addition, the mean daily value from the PSG nights will be summarized descriptively by study treatment and week. The mean value will be calculated using the same rule to calculate the primary endpoint (Section 4.2.1)), i.e., if there are data for both PSG nights, the mean value of the 2 nights will be used; if only 1 night is available, data from that one night will be used.

In addition, total compliance on the participant's use of the Sleepiz One+ device for the period of the treatment duration will be summarized using descriptive statistics by study treatment.

Compliance will be derived using the parameter good signal duration (GSD) which measures the daily total duration where the participant is in sight of the device. A participant will be compliant if the GSD has a total duration of at least 4 hours.

The compliance (as a percentage) will be calculated as

100 * number of days during treatment with GSD ≥ 4 hours / treatment duration

Treatment duration is defined in Section 4.5.1. If the participant has no Sleepiz One+ data during treatment, the participant will have a compliance of zero.

As well as a measure of compliance of the participant in using the device, GSD is also a measure of the quality of the data received and the subsequent processed metrics. Therefore, the descriptive statistics of the mean daily value and change from baseline in mean daily value for WASO, TST, SE and SOL will be repeated using only nights with GSD ≥ 4 hours when deriving the mean daily value.

4.6.4 Subgroup Analyses

Exploratory subgroup analyses using descriptive statistics will be provided for the primary and secondary endpoint change from baseline in WASO at Week 4 and 12 for the following subgroups:

- Average weekly frequency of moderate to severe HF over 14 days prior to respective visit confirmation date for the HF eligibility (<35 moderate to severe HF per week, ≥35 – <50 moderate to severe HF per week, ≥50 moderate to severe HF per week) - per actual stratification.
- Region (North America, Europe)
- BMI (<18.5, 18.5 to <25, 25 to <30, ≥ 30 kg/m^2)
- WASO as measured by PSG at baseline (quartiles)

If the number of participants is too small (less than 10%) within a subgroup, then the subgroup categories may be redefined prior to unblinding the study.

4.7 Interim Analyses

No interim analysis is planned.

4.8 Changes to Protocol-planned Analyses

In the protocol, Table 1-1 and Table 3-2 specifies that for the ICEs 'temporary study treatment interruption' and 'Intake of prohibited concomitant medication having impact on efficacy' data collected after the ICE will be discarded for the primary estimand. In the SAP, this is updated to clarify that only data collected during the ICE will be discarded (Section 1.1). If an ICE 'temporary study treatment interruption' and/or 'Intake of prohibited concomitant medication having impact on efficacy' occurred during the week 4 timeframe but not for week 12, the updated data handling method would not discard the week 12 assessment.

In addition, the definition of the ICE temporary study treatment interruption at week 4 in Section 9.3.1 of the protocol considers week 1 to week 4 for the calculation of treatment taken <80%. The definition was updated to consider week 0 in the calculation as well as including the PSG nights. The 'treatment taken on <5/7 days during either week 3 or 4' was also updated to consider the 7 days prior to the last PSG assessment since the visit may not be performed at the end of week 4. The same update to ICE temporary study treatment interruption at week 12 was implemented.

The MMRM model in Section 9.3.2.2. of the protocol specified that an interaction term baseline-by-week would be included as a covariate in the model. Due to the short treatment period and small sample size of the study, the baseline-by-week interaction term has been removed from the analytical model specified in Section 4.2.2 of the SAP.

A tipping point analysis was specified in the protocol as a sensitivity analysis for the primary estimand. This analysis has been removed from the SAP since this is an estimation study and a meaningful difference has not been established for this population to be used as the threshold to find the tipping point.

5. Sample Size Determination

The primary objective for the study is to estimate the treatment effect of elinzanetant compared to placebo in the change from baseline in WASO to week 4. Therefore, the sample size is based on precision (i.e. the half width of the confidence interval).

There are no data available on the primary endpoint WASO measured by PSG in women with sleep disturbance associated with the menopause. Therefore, the standard deviation used in the sample size calculation is obtained from a study in patients with insomnia disorder (Mignot et al. 2022). Assuming the pooled standard deviation is 33.2 for the change from baseline in WASO at week 4, a total sample size of 70 participants (35 per arm) would give a 90% probability that the 2-sided 90% confidence interval will have a half width of 14.7 minutes or less.

To account for missing data and possible inflation in variability due to missing data imputation, a total of approximately 78 (39 per treatment) will be randomized in a 1:1 ratio to both treatments. Approximately 223 participants will be screened to achieve 78 randomized participants.

6. Supporting Documentation

6.1 APPENDIX 1: Population Characteristic

6.1.1 Participant Disposition

The number of participants enrolled, randomized and valid for the FAS, SAF and SLAS will be summarized overall and by study treatment and study site.

The number of participants randomized will further be summarized by country/region and by study treatment and overall.

The number of sites will be summarized by country/region.

The number of participants with important protocol deviations and the number of screen failures will be summarized overall and by study treatment, country/region and study site.

The number of participants per important deviation category and validity finding will be presented by study treatment and overall.

A disposition summary will be presented summarizing the number of participants starting and completing each study period (screening, treatment, follow-up), the number of participants discontinuing it and the primary reason for discontinuation. This will be presented overall and for each study treatment. The number of participants completing the overall study will be provided by study arm and overall. Participants who completed the study are those who complete study treatment and were followed for the duration of the study. Follow-up is considered to be completed if the participants complete the follow-up visit.

6.1.2 Demography and Other Baseline Characteristics

All demographic and baseline characteristics will be summarized using descriptive statistics by study treatment and overall based on the FAS and SLAS.

Demographic and baseline assessments to be summarized will include:

- Age (at inclusion), race, ethnicity
- Age groups (<40 years, 40 - 49 years, 50 - 59 years, 60 - 65 years, > 65 years)
- Weight (kg), height (cm), body mass index (BMI; kg/m^2)
- BMI group (< 18.5, 18.5 to < 25, 25 to < 30, ≥ 30 kg/m^2)
- Smoking history (Never, Former, Current)
- Alcohol substance (abstinent, light, moderate, heavy)
- Caffeine consumption (mg) per day

Daily caffeine consumption will be calculated using data on the Caffeine Consumption eCRF page. The page captures the volume consumed (over a month, week, or day) for each type of caffeine substance (caffeinated coffee, caffeinated tea, energy drink, cola-type beverage, other). The caffeine content for each substance type in the below table will be used as a guide (Mayo Clinic).

Table 6-1: Caffeine Content

Substance type	Caffeine amount (mg) per 8 US fl oz
Caffeinated coffee	96
Caffeinated tea	47
Energy drink	72
Cola-type beverage	22
Other	22

For each 'Type of substance' on the eCRF page, follow the below steps:

1. Convert 'Average volume / amount consumed' into a ratio of 8 oz using the conversion table below:

'Unit of average volume / amount consumed'	Conversion factor
oz	oz / 8
mL	(mL x 0.034) / 8
dL	(dL x 3.381) / 8
L	(L x 33.814) / 8

2. Multiple by the caffeine amount specified in Table 6-1.
3. If 'Per time period' is recorded as 'Every Month', divide by 30.4. Else, if 'Per time period' is recorded as 'Every Week', divide by 7.

To calculate the caffeine consumption per day, sum the amount of caffeine for all types of substance together from step 3.

6.1.3 Reproductive and Menstrual History

Reproductive and menstrual history will include information on number of pregnancies, number of births, years being amenorrheic and number of participants with oophorectomy. Oophorectomy is based on Relevant Prior and Concomitant Procedures and the PTs Hysterosalpingo-oophorectomy, Oophorectomy, Oophorectomy bilateral, Salpingo-oophorectomy, Salpingo-oophorectomy bilateral, Salpingo-oophorectomy unilateral will be used. If the start date of the procedure is incomplete, the following rules will be applied to determine if the procedure is prior:

- If day is missing but the month and year is known, procedure will be prior if the month and year is before the month and year of start of treatment.
- If day and month are missing but year is known, procedure will be prior if the year is before the year of start of treatment.

These variables will be analyzed descriptively, separately for each treatment group and overall based on the SAF.

6.1.4 Protocol Deviations

Important deviations from the protocol and validity findings and the resulting assignment of participants to the analysis sets (see Section 3) are agreed upon in the blind review meeting (BRM). The documentation of important deviations, validity findings and the assignment of participant data to analysis sets will be performed according to the sponsor's applicable Standard Operating Procedures and/or Instruction Manuals. The definition for important deviations and validity findings will be provided in the 'Specification of assessment criteria and identification requirements' before unblinding the data.

Identification of important deviations and validity findings will be done periodically while the study is running, concluding with the completion of the final list during the BRM. Any changes to the statistical analysis prompted by the results of BRM will be documented in an amendment and, if applicable, in a supplement to this SAP.

The list of important deviations and validity findings will only be final after assessment of conditional findings, which are validity findings that can only be identified after unblinding of study treatment.

6.1.5 Medical History

Medical history will be tabulated using medical dictionary for regulatory activities (MedDRA; current version at the time of analysis) terms. The number of participants with medical history findings will be presented by primary system organ class and preferred term overall and per study treatment.

6.1.6 Relevant Prior and Concomitant Procedures

Relevant prior and concomitant procedures performed during the study other than those scheduled in the protocol will be tabulated using medical dictionary for regulatory activities (MedDRA; current version at the time of analysis) terms. The number of participants with relevant prior and concomitant procedures findings will be presented by primary system organ class and preferred term overall and per study treatment.

6.1.7 Prior, Concomitant and Post-Treatment Medication

For prior and concomitant medications, the following definitions apply:

- Prior medication: Medication taken before start of the study intervention intake, (regardless of when it ended).
- Concomitant medication: Medication taken during treatment period, i.e., between first and last study intervention intake (regardless of when it started or ended).
- Post-treatment medication: Start of medication is after last study drug intake.

Missing dates will be imputed as described in Section 4.1.1.4.

Prior, concomitant and post-treatment medication will be coded to Anatomical Therapeutic Chemical (ATC) classification codes according to the World Health Organization Drug Dictionary (WHO-DD, current version at the time of analysis). The number of participants taking prior, concomitant, post-treatment medication, and prohibited concomitant medications will be presented by study treatment and overall using ATC classes and subclasses.

6.2 APPENDIX 2: Definitions of Sleep Parameters

6.2.1 PSG Endpoints

PSG recordings will be sent to a vendor for central scoring. Files will be scored manually, and sleep parameters will be provided using the definitions in the below table:

Table 6-2: PSG Endpoint Definitions

Parameter	Definition
Wakefulness after sleep onset (WASO)	Minutes of wake from onset of persistent sleep to lights on. Persistent sleep defined as 20 consecutive epochs (10 minutes) on non-wakefulness.
Sleep efficiency (SE)	(Total sleep time / time in bed) x 100.
Latency to persistent sleep (LPS)	Duration in minutes from lights off to the first epoch of 20 consecutive epochs of non-wakefulness.
Total sleep time (TST)	Duration of total REM and NREM sleep during time in bed.
Sleep onset latency to N1 (SOLN1)	Duration in minutes from lights off to the first epoch of non-wakefulness (N1, N2, N3, or REM).
Sleep onset latency (SOL)	Duration in minutes from lights off to the first epoch of sleep other than N1 (i.e., N2, N3, or REM).
Wake time during sleep (WTDS)	Minutes of wake from onset of persistent sleep to the last epoch of sleep stage N2, N3 or REM before lights on. That is, does not include "terminal wake" from last epoch of N2, N3, or REM before lights on.
Early morning wake time (EMWT)	Minutes of wake from the last epoch of sleep stage N2, N3 or REM prior lights on to lights on or end of the file.
Total wake time (TWT)	Total wake time from lights off to lights on.
Percentage of TWT	(TWT / time in bed) x 100.
Number of awakenings (NAW)	Count total number of awakenings from onset of persistent sleep to lights on. An awakening is defined as a PSG recording of at least two consecutive wake epochs. An awakening must be separated by an epoch of stage N2 or stage N3 or stage R. Separation by stage N1 does not denote a new awakening. Terminal awakening will not be counted. Terminal awakening is the last awakening prior to lights-on and the participant remains awake until lights-on. If participant is asleep at lights-on, there is no terminal awakening.
Mean duration of awakenings	Average of all awakenings after onset of persistent sleep, not including any terminal awakening.
WASO by quarter	Minutes of wake for each quarter, starting from lights off, where each quarter is 2 hours long. If total recording is not 8 hours, then some quarter may not be available. If persistent sleep is not achieved by that quarter, then quarter will be reported blank.

Stage N1 sleep time	Total duration in minutes in stage N1 from lights off to lights on.
Percentage of stage N1 sleep	(Stage N1 sleep time / TST) x 100.
Stage N2 sleep time	Total duration in minutes in stage N2 from lights off to lights on.
Percentage of stage N2 sleep	(Stage N2 sleep time / TST) x 100.
Stage N3 sleep time	Total duration in minutes in stage N3 from lights off to lights on.
Percentage of stage N3 sleep	(Stage N3 sleep time / TST) x 100.
Non-REM sleep time	Total duration in minutes in non-REM sleep (Stage N1+N2+N3) from lights off to lights on.
Percentage of Non-REM sleep	(Non-REM sleep time / TST) x 100.
REM sleep time	Total duration in minutes in stage REM from lights off to lights on.
Percentage of REM sleep	(REM sleep time / TST) x 100.
REM onset latency from sleep onset	Duration from sleep onset (first epoch of sleep) to the first REM epoch.
Periodic limb movement disorders with arousal index (PLMAI)	Number of PLM with arousal per hour during total sleep time.
Apnea hypopnea index (AHI)	Total count of Apnea and Hypopnea per hour during total sleep time.
Arousal index	An arousal is an abrupt shift in EEG frequency, which may include theta, alpha, and/or frequencies greater than 16 Hz but not spindles. The minimum duration of an arousal event is 3 seconds, with at least 10 seconds of stable sleep preceding the change. Scoring arousals during REM must also be accompanied by a concurrent increase in submental EMG lasting at least 1 second. The index is based on the number of all arousals scored and counted during sleep time reported per hour.
Number of stage shifts	Number of stage shifts from stage 2, 3, or REM to stage 1 or Wake from lights off to lights on.

6.2.2 Sleepiz One+ Endpoints

The Sleepiz One+ monitoring device will capture and record raw data, which will be processed by the vendor's software to derive the nightly parameters as defined below:

Table 6-3: Sleepiz One+ Endpoint Definitions

Parameter	Definition
Wakefulness after sleep onset (WASO)	Time duration, in minutes, the monitored subject was awake between the sleep onset (first sleep period longer than 10 minutes) and the last awakening.
Sleep efficiency (SE)	Percentage of time the monitored subject is asleep while in bed, from the person's last entry into bed before falling asleep, and the person's last exit from the bed.
Total sleep time (TST)	The total time during the recording where the subject was asleep.
Sleep onset latency (SOL)	Time duration, in minutes, from the person's last entry into bed before falling asleep, to the first sleep event with a duration greater or equal to 10 minutes.
Wake up latency	Time duration, in minutes, between the last transition from asleep to awake to the person exiting the bed for 10 minutes or more.
Awakening count	Number of times the monitored subject woke up between the sleep onset (first sleep period longer than 10 minutes) and the last awakening.
Average awakenings	The average duration of awakenings (in mins)
Out of bed count	Number of times the monitored subject was out of bed (no presence detected) between the sleep onset (first sleep period longer than 10 minutes) and the last awakening.
Time in bed (TIB)	Total time during the recording where the subject was in bed.
Respiratory irregularity index (RII)	The total number of abnormal breathing events (pauses) per hour of sleep, measured via radar. This value is analogous to the apnea-hypopnea index (AHI).
Heart rate per minute	The minimum, maximum, mean, median,10th and 90th percentile heart rate over the night.
Breathing rate per minute	The minimum, maximum, mean, median,10th and 90th percentile breathing rate over the night.
Good signal duration	Time duration where the subject is in sight of the device and where either movement, breathing rate, or heart rate can be computed.

6.3 APPENDIX 3: Identification of Prohibited Medications

Drug groupings, individual drug names and manual review will be used to identify prohibited therapies and medications that were defined in protocol Section 10.7. This includes prior hormonal therapies, medications that affect PK, and medications potentially confounding efficacy that are to be washed out during the screening period as well as the identification of the ICE prohibited concomitant medications impacting efficacy.

The list of prohibited medications together with the pre-defined washout time period for their effect will be reviewed during blind review of the data and finalized prior to unblinding.

Table 6-4: Preliminary list of prohibited hormonal therapies by drug groupings

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Impact duration of concomitant medication
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	109	GnRH agonists	From 1st dose to 12 weeks after the last dose.
		108	Drugs acting on gonadotropin-releasing hormone (GnRH) receptors	110	GnRH antagonists	From 1st dose to 12 weeks after the last dose.
		5	Hormone replacement therapy	2	Oestrogens used in replacement therapy	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal, topical or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral or sublingual. 3, From 1st dose to 12 weeks after the last dose if the route is intramuscular. 4, From 1st dose to 24 weeks after the last dose if the route is subcutaneous or intradermal.
		5	Hormone replacement therapy	3	Progestogens used in replacement therapy	1, From 1st dose to 4 weeks after the last dose if the route is vaginal, intravaginal, cutaneous, nasal, topical or transdermal. 2, From 1st dose to 8 weeks after the last dose if the route is oral or sublingual. 3, From 1st dose to 12 weeks after the last dose if the route is subcutaneous or intradermal. 4, From 1st dose to 24 weeks after the last dose if the route is intramuscular.

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Impact duration of concomitant medication
		772	Cancer therapies	738	Endocrine antineoplastic therapy	From 1st dose to 12 weeks after the last dose.

Table 6-5: Preliminary list of prohibited concomitant medication impacting PK by drug grouping

DGCODEL0	DGNAME0	DGCODEL1	DGNAME1	DGCODEL2	DGNAME2	Impact duration of concomitant medication
45	Drugs interacting with CYP3A	240	CYP3A inducers	225	Moderate CYP3A inducers	From 1st dose to 4 weeks after the last dose Considered not to have an influence on efficacy, if route is auricular(otic), ophthalmic or conjunctival.
45	Drugs interacting with CYP3A	240	CYP3A inducers	265	Strong CYP3A inducers	From 1st dose to 4 weeks after the last dose

All listed drug names in Table 6-6 were reported in NIRVANA as of finalization of this SAP.

Table 6-6: Preliminary list of prohibited concomitant medication potentially confounding efficacy by drug names

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	Impact duration of concomitant medication
DIPHENHYDRAMINE HYDROCHLORIDE	000004	02	From 1st dose to 2 weeks after the last dose
BETAMETHASONE DIPROPIONATE;BETAMETHASONE SODIUM PHOSPHATE	000085	13	From 1st dose to 2 weeks after the last dose if route is oral.
DIAZEPAM	000170	01	From 1st dose to 4 weeks after the last dose
METHOCARBAMOL	000479	01	From 1st dose to 2 weeks after the last dose
METHYLPREDNISOLONE	000496	01	From 1st dose to 2 weeks after the last dose if route is oral
HYDROXYZINE	000584	01	From 1st dose to 2 weeks after the last dose
HYDROXYZINE HYDROCHLORIDE	000584	02	From 1st dose to 2 weeks after the last dose

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	Impact duration of concomitant medication
CLONIDINE	001711	01	From 1st dose to 2 weeks after the last dose
CLONAZEPAM	002852	01	From 1st dose to 4 weeks after the last dose
CYCLOBENZAPRINE	004284	01	From 1st dose to 2 weeks after the last dose
CITALOPRAM	005826	01	From 1st dose to 2 weeks after the last dose
CITALOPRAM HYDROBROMIDE	005826	02	From 1st dose to 2 weeks after the last dose
TRAMADOL	005992	01	From 1st dose to 2 weeks after the last dose
TRAMADOL HYDROCHLORIDE	005992	02	From 1st dose to 2 weeks after the last dose
BUPROPION HYDROCHLORIDE	007005	02	From 1st dose to 2 weeks after the last dose
TIZANIDINE HYDROCHLORIDE	007407	02	From 1st dose to 2 weeks after the last dose
CETIRIZINE HYDROCHLORIDE *	008843	02	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
LORATADINE *	009175	01	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
GABAPENTIN	010030	01	From 1st dose to 2 weeks after the last dose
TOPIRAMATE	012017	01	From 1st dose to 2 weeks after the last dose
VENLAFAXINE HYDROCHLORIDE	012338	02	From 1st dose to 2 weeks after the last dose
MELATONIN	012432	01	From 1st dose to 2 weeks after the last dose
FEXOFENADINE HYDROCHLORIDE *	013142	02	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
FEXOFENADINE;PSEUDOEPHEDRINE HYDROCHLORIDE *	129590	03	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
DESLORATADINE *	013985	01	From 1st dose to 2 weeks after the last dose if

WHO Drug Name	WHO Drug Record Number	WHO Drug Sequence Number 1	Impact duration of concomitant medication
			frequency is once daily (QD) or 3 times per week
LEVOCETIRIZINE DIHYDROCHLORIDE *	015302	02	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
BILASTINE *	071040	01	From 1st dose to 2 weeks after the last dose if frequency is once daily (QD) or 3 times per week
ACTAEA RACEMOSA **	014568	01	From 1st dose to 2 weeks after the last dose
ACTAEA RACEMOSA EXTRACT **	014568	04	From 1st dose to 2 weeks after the last dose
ACTAEA RACEMOSA EXTRACT;BORON;CALCIUM;FOLIC ACID;GLYCINE MAX EXTRACT;MAGNOLIA OFFICINALIS;NICOTINIC ACID;PYRIDOXINE HYDROCHLORIDE;RIBOFLAVIN;SELENIUM;THIAMINE;TOCOPHEROL;VITAMIN B12 NOS **	113355	02	From 1st dose to 2 weeks after the last dose
DOXYLAMINE SUCCINATE	003341	02	From 1st dose to 2 weeks after the last dose
FEZOLINETANT	093415	01	From 1st dose to 2 weeks after the last dose
HYDROCODONE	000600	01	From 1st dose to 2 weeks after the last dose
ALPRAZOLAM	005952	01	From 1st dose to 2 weeks after the last dose
OXYBUTYNIN	005389	01	From 1st dose to 2 weeks after the last dose
NIRMATRELVIR;RITONAVIR	158921	01	From 1st dose to 2 weeks after the last dose
TRIAMCINOLONE ACETONIDE	000319	02	From 1st dose to 2 weeks after the last dose
PHENTERMINE	001317	01	From 1st dose to 2 weeks after the last dose

* Non-sedating antihistamines/pseudoephedrine are allowed up to twice per week per protocol. Participants with PRN or ‘as needed’ are considered as taking these medications according to the study protocol.

**These medications belong in the drug grouping DGNAME2 ‘Oestrogen used in replacement therapy’ and should have impact duration specified in this table instead of the impact duration specified in Table 6-4.

6.4 APPENDIX 4: Identification of Intercurrent Events

6.4.1 Premature discontinuation of study treatment

Participants who prematurely discontinue study treatment also withdrawal from study participation at the same time. No further study procedures (except for the safety follow-up visit) will be taken including PSG assessments.

Participants will likely not have a PSG assessment after discontinue study treatment. Therefore, for the identification of the ICE premature discontinuation of study treatment at week 4 and week 12, visit day 30 and day 86 will be used as the PSG assessment day respectively.

6.4.2 Temporary treatment interruption

Temporary treatment interruption is defined as

- Treatment taken < 80% during day 1 to the last PSG night at week 4 (inclusive) OR treatment taken on < 5 days during the last 7 days up to the last PSG night at week 4 (inclusive) for week 4.
- Treatment taken < 80% during day 1 to the last PSG night at week 12 (inclusive) OR treatment taken on < 5 days during the last 7 days up to the last PSG night at week 12 (inclusive) for week 12.

If the week 4 PSG assessment is missing, day 30 will be used instead of the last PSG night in the definition of treatment interruption for week 4 only for participants who have not prematurely discontinue treatment prior to day 30. Similarly for week 12, day 86 will be used only for participants who have not prematurely discontinue treatment prior to day 86.

The criterion for treatment taken on 5/7 days during the corresponding week prior to and including the last PSG night, as defined above, is met, if treatment has been taken for less than 5 days (i.e., at least 3 missed intakes). To calculate the number of days per week with treatment intake, a day where at least one capsule was taken, will be considered as a day with treatment intake.

To identify whether < 80% treatment was taken, compliance up to and including the last PSG night at week 4 will be calculated as

100 * Number of capsules taken up to and including the last PSG night at week 4 /
(2 * study duration up to and including the last PSG night at week 4)

Compliance up to and including the last PSG night at week 12 will be calculated as

100 * Number of capsules taken up to and including the last PSG night at week 12 /
(2 * study duration up to and including the last PSG night at week 12)

Study duration up to and including the last PSG night at week 4 and week 12 is derived as the number of days from the day of first study intervention intake up to and including the last PSG night at week 4 and week 12 respectively. The number of capsules taken up to and including the last PSG night at week 4 and week 12 will be calculated based on the ePRO instrument ‘Study medication intake documentation’. If the ‘Study medication intake documentation’ on the second PSG night at week 12/EoT is missing but the last treatment intake on the eCRF ‘Study intervention’ pages indicate that the participant took treatment on

the night, the ‘Number of study medication capsules taken’ on the ‘Study medication intake documentation’ will be imputed with 1 capsule.

6.4.3 Intake of prohibited concomitant medication having impact on efficacy

The list of prohibited concomitant medications that were defined in the protocol to influence efficacy will be finalized prior to unblinding. Missing dates will be imputed as described in Section 4.1.1.4.

PSG assessments during any intake of prohibited medication or intake that occurred prior to the PSG assessment and is within the washout period of the prohibited medication (i.e., within the timeframe specified in the impact duration column of Table 6-4, Table 6-5and Table 6-6) will be considered as an intercurrent event. PSG assessments will be performed on 2 consecutive nights at each visit and if only one of the nights fall within the timeframe, then that night will be discarded while the other night will be included in the analysis. For the identification of this ICE, visit day 30 and day 86 will be used as the PSG assessment day for week 4 and week 12 respectively if PSG assessment(s) are missing. The preliminary list of prohibited concomitant medications are specified in Section 6.3.

6.5 APPENDIX 5: Coding Conditions Applicable for AESI

AESI	Search
Potential AESI -Liver event "Any condition triggering close liver observation" according to protocol section 10.5 results in true AESIs of liver events.	SMQ Cholestasis and jaundice of hepatic organ SMQ Drug related hepatic disorders – severe events only SMQ Liver related investigations, signs and symptoms SMQ Liver-related coagulation and bleeding disturbances In addition, include AESIs ticked at AE eCRF.
Post-menopausal bleeding	MLG Female genital track bleeding PT Abnormal uterine bleeding PT Abnormal withdrawal bleeding PT Cervix haematoma uterine PT Cervix haemorrhage uterine PT Coital bleeding PT Haematocoele female PT Haematosalpinx PT Haemorrhagic ovarian cyst PT Ovarian haematoma PT Ovarian haemorrhage PT postmenopausal haemorrhage PT Unexpected vaginal bleeding on hormonal IUD PT Uterine haematoma PT Uterine haemorrhage PT Vaginal haematoma PT Vaginal haemorrhage PT Vulval haematoma PT Vulval haematoma evacuation PT Withdrawal bleed

7. References

Baker FC, de Zambotti M, Colrain IM, Bei B. Sleep problems during the menopausal transition: prevalence, impact, and management challenges. Nat Sci Sleep. 2018a;10:73-95.

Baker FC, Lampio L, Saaresranta T, Polo-Kantola P. Sleep and Sleep Disorders in the Menopausal Transition. Sleep Med Clin. 2018b Sep;13(3):443-56.

Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2001;2(4):297-307.

Coborn J, de Wit A, Crawford S, Nathan M, Rahman S, Finkelstein L, et al. Disruption of Sleep Continuity During the Perimenopause: Associations with Female Reproductive Hormone Profiles. J Clin Endocrinol Metab. 2022 Sep 28;107(10):e4144-e53.

FDA. Patient-focused drug development guidance public workshop. Methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. 2018.

Guthrie KA, Larson JC, Ensrud KE, Anderson GL, Carpenter JS, Freeman EW, et al. Effects of Pharmacologic and Nonpharmacologic Interventions on Insomnia Symptoms and Self-reported Sleep Quality in Women With Hot Flashes: A Pooled Analysis of Individual Participant Data From Four MsFLASH Trials. Sleep. 2018 Jan 1;41(1).

Lieb K, Ahlvers K, Dancker K, Strohbusch S, Reincke M, Feige B, et al. Effects of the neuropeptide substance P on sleep, mood, and neuroendocrine measures in healthy young men. Neuropsychopharmacology. 2002 Dec;27(6):1041-9.

Mayo Clinic. Caffeine content for coffee, tea, soda and more. Available online: https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/caffeine/art-20049372Accessed: 21 Nov 2023

Mignot E, Mayleben D, Fietze I, Leger D, Zammit G, Bassetti CLA, et al. Safety and efficacy of daridorexant in patients with insomnia disorder: results from two multicentre, randomised, double-blind, placebo-controlled, phase 3 trials. Lancet Neurol. 2022 Feb;21(2):125-39.

Pengo MF, Won CH, Bourjeily G. Sleep in Women Across the Life Span. Chest. 2018 Jul;154(1):196-206.

PROMIS Sleep Disturbance – Scoring Manual. Available online: Accessed: 10 Aug 2023.

Ratti E, Carpenter DJ, Zamuner S, Fernandes S, Squassante L, Danker-Hopfe H, et al. Efficacy of vestipitant, a neurokinin-1 receptor antagonist, in primary insomnia. Sleep. 2013 Dec 1;36(12):1823-30.

Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.

Simon JA, Anderson RA, Ballantyne E, Bolognese J, Caetano C, Joffe H, et al. Efficacy and safety of elinzanetant, a selective neurokinin-1,3 receptor antagonist for vasomotor symptoms: a dose-finding clinical trial (SWITCH-1). Menopause. 2023 Jan 31;30(3).

Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006 May 22;166(10):1092-7.

Wang Y, Tu W, Kim Y, Sinks S, He J, Cambon A, et al. Statistical methods for handling missing data to align with threatment policy. Pharma Stats. 2023 July/August; 22 (4): 650-670.

Woods NF, Hohensee C, Carpenter JS, Cohen L, Ensrud K, Freeman EW, et al. Symptom clusters among MsFLASH clinical trial participants. Menopause. 2016 Feb;23(2):158-65.

Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS sleep disturbance and Sleep-Related Impairment item banks. Behav Sleep Med. 2011 Dec 28;10(1):6-24.