| | |
|---|---|
| **Document Type:** | Statistical Analysis Plan |
| **Official Title:** | A Double-Blind, Randomised, Placebo Controlled, Adaptive Design Study of the Efficacy, Safety and Pharmacokinetics of NT-814 in Female Subjects With Moderate to Severe Vasomotor Symptoms Associated With the Menopause |
| **NCT Number:** | NCT03596762 |
| **Document Date:** | 11 DEC 2019 |

# STATISTICAL ANALYSIS PLAN

## Protocol 814-PM-02

## A double-blind, randomized, placebo controlled, adaptive design study of the efficacy, safety and pharmacokinetics of NT-814 in female subjects with moderate to severe vasomotor symptoms associated with the menopause

| | |
|---|---|
| **Protocol Number:**<br>**(Version Date)** | 814-PM-02<br>Version 2 dated 07 February 2019 |
| **Name of Test Drug:** | NT-814 |
| **Phase:** | 2b |
| **Methodology:** | Randomized, Double-blind, placebo controlled, adaptive |
| **Sponsor:** | NeRRe Therapeutics Ltd,<br>SBC, Incubator Building,<br>Gunnels Wood Rd.,<br>Stevenage,<br>SG1 2FX<br>United Kingdom |
| **Sponsor Representative:** | PPD<br>NeRRe Therapeutics Ltd<br>SBC, Incubator Building,<br>Gunnels Wood Rd.,<br>Stevenage, SG1 2FX, UK<br>Phone: PPD<br>Email: PPD |
| **Document Date:** | 11 December 2019 |
| **Document Version:** | Final Version 3.0 |

## SIGNATURE PAGE

**Protocol Title:** A double-blind, randomized, placebo controlled, adaptive design study of the efficacy, safety and pharmacokinetics of NT-814 in female subjects with moderate to severe vasomotor symptoms associated with the menopause

**Sponsor:** NeRRe Therapeutics Ltd,

SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage,
SG1 2FX
United Kingdom

**Protocol Number:** 814-PM-02

**Document Date/Version:** 11 December 2019 / Final Version 3.0

**Cytel, Inc. Author:**
PPD
Cytel Inc. Geneva Branch|
20 Route de Pre-Bois
CH-1216 Cointrin

Signature: PPD Electronically signed by PPD
Reason: Approve
Date: 2019-12-11 09:07:55-05:00

Date: 11-Dec-2019

### Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol, and all applicable regulatory guidances and guidelines.

I have discussed any questions I have regarding the contents of this document with the biostatistical author.

I also understand that any subsequent changes to the planned statistical analyses, as described herein, may have a regulatory impact and/or result in timeline adjustments. All changes to the planned analyses will be described in the clinical study report (CSR).

**Sponsor Signatory:**
PPD
NeRRe Therapeutics Ltd
SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage, SG1 2FX, UK

Signature: PPD Electronically signed by PPD
Reason: Approve
Date: 2019-12-11 16:29:09Z

Date: 11-Dec-2019

# TABLE OF CONTENTS

**Section** **Page**

# LIST OF IN-TEXT TABLES

## ABBREVIATIONS

| Abbreviation | Definition |
|---|---|
| AE | Adverse Event |
| ALT | Alanine amino transferase / alanine transaminase |
| AST | Aspartate amino transferase / aspartate transaminase |
| ATC | Anatomic Therapeutic Class |
| BALP | Bone-specific alkaline phosphatase |
| BDI-II | Beck Depression Inventory – Version II |
| BDRM | Blinded Data Review Meeting |
| BMI | Body Mass Index |
| eCRF | Electronic Case Report Form |
| CSR | Clinical Study Report |
| eC-SSRS | Electronic Columbia Suicide Severity Rating Scale |
| CSR | Clinical Study Report |
| DRC | Data Review Committee |
| ECG | Electrocardiogram |
| EDC | Electronic Data Capture |
| eDiary | Electronic Diary |
| FAS | Full Analysis Set |
| GGT | Gamma Glutamyl Transferase |
| HFRDIS | Hot Flash Related Daily Interference Scale |
| HF | Hot Flush / Hot Flash |
| IA | Interim Analysis |
| IC | Informed Consent |
| ICH | International Conference for Harmonization |
| IEC | Independent Ethics Committee |
| IMP | Investigational Medicinal Product |
| IRB | Institutional Review Board |
| ISC | Independent Statistical Centre |
| ISI | Insomnia Severity Index |
| IWRS | Interactive Web Response Services |
| KaNDy | KaNDy Therapeutics Ltd |
| LOCF | Last Observation Carried Forward |
| MCV | Mean Corpuscular Volume |
| MeDRA | Medical Dictionary for Regulatory Activities |
| MenQoL-I | Menopause-specific Quality-of-Life Questionnaire Intervention Version |
| MMRM | Mixed-effect Model Repeated Measures |
| NeRRe | NeRRe Therapeutics Ltd |
| NTA | Night-time awakening |

| Abbreviation | Definition |
|---|---|
| P1NP | Procollagen type 1 N-terminal pro-peptide |
| PK | Pharmacokinetic |
| PP | Per-Protocol |
| PSQI | Pittsburgh Sleep Quality Index |
| PT | Preferred Term |
| RBC | Red Blood Cell |
| REML | Restricted Maximum Likelihood |
| SAE | Serious Adverse Event |
| SAP | Statistical Analysis Plan |
| SAS | Statistical Analysis System |
| SD | Standard Deviation |
| SOC | System Organ Class |
| SUSAR | Suspected Unexpected Serious Adverse Reaction |
| TEAE | Treatment Emergent Adverse Event |
| TMF | Trial Master File |
| WBC | White Blood Cell |
| WHO | World Health Organization |

# 1. INTRODUCTION AND OBJECTIVES OF ANALYSIS

## 1.1. Introduction

The study is a dose-range finding study, with assessment of dose-response based on both efficacy and safety. A four-fold range of doses will be evaluated.

The study will have an adaptive design component in which the number of subjects randomised to each treatment group is modified on the basis of emerging safety and efficacy data. This design enables a broad range of doses to be evaluated in the first instance, but with the ability to reduce the number of doses of NT-814 being evaluated if this emerging data shows that there is no advantage in continuing to evaluate them all. The study will comprise a 3-week screening and baseline period followed by a 12-week treatment period then a 4-week follow-up period.

This document refers to Protocol No. 814-PM-02 version 2.0, 7 February 2019, and to the annotated CRF Final Version 5.0 dated 15 July 2019.

The SAP is prepared in compliance with the ICH E9 Guideline on Statistical Principles for Clinical trials.

## 1.2. Objectives of Statistical Analysis

This statistical analysis plan (SAP) outlines the statistical methods, data derivations and summaries to be used in the interim (IA) and end of study (Week 16) analyses of study data in order to answer the study objectives. Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided. The statistical analyses and summary tabulations described in this SAP will provide the basis for the results sections of the clinical study report (CSR) for this trial.

Dedicated sections in this SAP will highlight the statistical analyses and summary tabulations to be performed in the interim analyses and in the final analysis, separately.

This SAP will also outline any differences in the currently planned analytical objectives relative to those outlined in the study protocol.

## 1.3. Team involved in the study

Two separate teams will be involved in this study:

- An independent statistical centre (ISC) composed of an unblinded statistician and programmers. They will be in charge of preparing outputs for the interim analyses. These individuals will work out of an office that is in a separate geographic location to the sponsor and blinded team.
- A blinded team composed of a blinded statistician and programmers. They will be responsible of preparing the outputs for the end of study analysis.

For further details, the charter describes the operation of the interim reviews, as managed by the Data Review Committee (DRC).

# 2. STUDY DESIGN

## 2.1. Synopsis of Study Design

This is a multi-centre, multi-country, double-blind, randomised, placebo-controlled Phase 2b study. The study will have a single-blind placebo run-in period and will be adaptive with respect to the number of subjects recruited into each dose group.

Four doses of NT-814 (40 mg once a day, 80 mg once a day, 120 mg once a day and 160 mg once a day) will be investigated and compared to placebo, in five parallel groups. All subjects will receive placebo for the last 2 weeks of the screening/baseline period, after which subjects who meet all of the eligibility criteria will be randomised into the study. Subjects will initially be randomised 1:1:1:1:1 to each of the treatment groups, with the randomisation ratio subject to change in response to emerging efficacy and safety data. Regardless of any changes to the randomisation ratio, subjects will continue to receive the dose regimen they were randomised to for the full 12-week treatment period.

Subjects will participate in the study for a total of approximately 19 weeks, comprising a screening & baseline period (3 weeks), 12 weeks of double-blind treatment and then a final follow up visit, 4 weeks after the end of the treatment period. There will be a total of 8 visits whilst participating in the study.

## 2.2. Randomization Methodology

Eligible subjects will be randomised on Day 1 to NT-814 or placebo. Randomisation will be performed centrally via an interactive web response system (IWRS) and is integrated into the Electronic Data Capture (EDC) system. The randomisation schedule (both the initial and any updates following IA) will be produced and entered into the IWRS by Pharm Olam International according to their Standard Operating Procedures.

An adaptive design will be used to maximise the efficiency of the study through discontinuation or reduction in the number of subjects randomised to doses that are found, on the basis of emerging data, to be either insufficiently effective or no more effective than an adjacent lower dose. Doses may also be discontinued for safety reasons and the randomisation ratio may be adjusted in favour of one or more doses without necessarily discontinuing other doses.

On successful completion of screening and confirmation of eligibility, subjects will initially be randomised in a 1:1:1:1:1 ratio to receive either NT-814 40 mg/day, 80 mg/day, 120 mg/day, 160 mg/day or placebo. Randomisation will be stratified by region (North America or Europe). After completion of the first interim review the randomisation ratio may be changed. Regardless of any changes to the randomisation ratio, subjects will continue to receive the dose regimen they were randomised to for the full 12-week treatment period.

The initial estimated sample size for the study is 165 subjects. However, the total number of subjects may be higher or lower according to the outcome of the interim analyses.

## 2.3. Stopping Rules and Unblinding

Patients can decide at any time to withdraw consent from the study. Investigators can decide at any time during the study to discontinue treatment for an individual patient based on their own medical judgment. Reasons for discontinuing a subject should be documented in the CRF, and patients will be encouraged to participate to the final follow-up/early termination visit even if they discontinue study medication.

The Sponsor may terminate the study for safety or administrative reasons at any time. The Data Review Committee (DRC) is responsible for overseeing the safety of the study and the DRC Charter includes specific criteria for stopping or suspending the study.

Study investigators (principal and sub) will be given access to the IWRS system for the purposes of emergency unblinding. Investigators are permitted to unblind treatment for a subject if it is deemed that knowledge of the subject's treatment will impact subject's future medical care. If a suspected unexpected serious adverse reaction (SUSAR) occurs that requires expedited reporting to the relevant regulatory agency/Institutional Review Board (IRB) /Independent Ethics Committee (IEC), then the blind will be broken for the relevant subject by Emas safety group through IWRS, in order to provide the regulatory agencies with the knowledge of the event and the causal agent. A blinded (or an unblinded report if required) copy of the report will be provided to the investigators and the relevant IRB/IEC. The study operational team will remain blinded to treatment allocation until the database has been locked at the end of the study.

If unblinding occurs accidentally this will be considered a protocol deviation which must be documented in the subject's medical notes and in the trial master file (TMF), and the Sponsor must be informed.

## 2.4. Study Procedures

The schedule of assessments, as outlined in the study protocol, is provided in

Table 1.

## Table 1 Schedule of Assessments

| Procedure | Screening Visit 1 (Visit 1) | Screening Visit 2 (Visit 2) | Baseline (Visit 3) | Week 2 (Visit 4) | Week 4 (Visit 5) | Week 8 (Visit 6) | Week 12 (Visit 7) | Week 16 Follow-up/Early Termination visit[b] (Visit 8) |
|---|---|---|---|---|---|---|---|---|
| **Visit Day** | **-21[a]** | **-14** | **1** | **15** | **29** | **57** | **85** | **113** |
| **Allowable window** | **[a]** | **±2 days** | **0** | **±3 days** | **±3 days** | **±4 days** | **±4 days** | **±5 days** |
| Informed Consent[c] | X | | | | | | | |
| Medical History/Concomitant Diseases | X | X | | | | | | |
| Demography | X | | | | | | | |
| Physical Exam | | X | X[d] | | | | | X[d] |
| Inclusion/Exclusion Criteria | X | X[e] | X[f] | | | | | |
| Review of Concomitant Medications | X ---------- | | | | | | | ----------X |
| Vital Signs[g] | | X | X | X | X | X | X | X |
| 12-lead ECG | | X | X | X | X | X | X | X |
| AE and SAE Recording | | X ---------- | | | | | | ---------- X |
| Issue Paper Diary / Training | X | | | | | | | |
| Paper Diary Completion (once daily) | X-------X[h] | | | | | | | |
| Issue eDiary / Training / Compliance check | | X | X | X | X | X | X | |
| eDiary Completion (twice daily) | | X---------- | | | | | | ----------X[i] |
| Study Drug Dispensing/Training | | X | X | | X | X | | |
| Placebo Treatment | | X-------X[j] | | | | | | |
| Study Drug Collection/Compliance | | | X | X | X | X | X | |
| Randomisation | | | X | | | | | |
| Daily Dosing (evening before bedtime) | | | X ---------- | | | | X | |
| MenQoL-I | | | X | | X | X | X | X |
| HFRDIS | | | X | X | X | X | X | X |
| Pittsburgh Sleep Quality Index | | | X | | X | X | X | X |
| Insomnia Severity Index | | | X | | X | X | X | X |
| Columbia Suicide Severity Rating Scale | | | X | | X | | X | X |
| Beck Depression Inventory II | | | X | X | X | X | X | X |
| Clinical Chemistry and Haematology | | X | X | X | X | X[k] | X | X |
| Blood sample for NT-814 concentration | | | | X | X | X | X | |
| Blood sample for bone turnover markers | | | X | | | | X | X |
| Urine pregnancy Test | | X | | | | | | |
| Urinalysis | | X | X | X | X | | X | X |

[a] Screening visit 1 must be at least 6 days before Screening Visit 2 to allow at least 6 days of paper diary data to be completed. It can be earlier than Day -21 if needed to enable prohibited concomitant medications to be washed-out.
[b] An Early Withdrawal/Safety Follow-up visit will be performed if the subject withdraws after the Baseline visit. This visit will consist of the same assessments as the Week 16 Follow-up visit and should be scheduled within 14 days of the early termination date or discontinuation of investigational product.
[c] Informed consent will be obtained prior to any screening procedures being performed. Consent can be obtained prior to Screening Visit 1 or at the actual visit.
[d] Symptom directed examination, if required.
[e] The paper diary Hot Flush requirements, as well as the other inclusion/exclusion criteria, will be assessed at Screening Visit 2.
[f] The eDiary Hot Flush requirements will be assessed and other inclusion/exclusion criteria will be reviewed at the Baseline visit.
[g] Systolic and diastolic blood pressure, pulse rate, temperature and weight will be recorded at all visits except Screening Visit 1. Waist circumference will be recorded at all visits except Screening Visits 1 and 2. Height will be measured at Screening Visit 2 only.
[h] Subjects will complete a paper diary once a day for 7 days in between Screening Visit 1 and Screening Visit 2 (a minimum of 6 days of diary data is required to confirm eligibility).
[i] Subjects will complete the eDiary twice a day, from the evening of Screening Visit 2 until they exit the study.
[j] At Screening Visit 2, subjects will the commence placebo treatment (the placebo run-in period). Their first dose of placebo will be taken in the clinic and all subsequent doses at home once a day in the evening (starting the following day) up until the day before the Baseline visit.
[k] At Week 8, this blood sample is for clinical chemistry only.

## 2.5. Efficacy, Pharmacokinetic, and Safety Variables

### 2.5.1. Efficacy Variables

#### 2.5.1.1. Primary Efficacy Variables

There are four co-primary efficacy endpoints:

- Mean change from baseline in mean daily frequency of moderate and severe hot flushes (HF) from baseline to Week 4
- Mean change from baseline in mean daily frequency of moderate and severe HF from baseline to Week 12
- Mean change from baseline in mean severity of moderate and severe HF from baseline to Week 4
- Mean change from baseline in mean severity of moderate and severe HF from baseline to Week 12

#### 2.5.1.2. Secondary Efficacy Variables

The secondary efficacy endpoints include:

- Mean change from baseline in frequency of mean daily moderate and severe HF from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in mean severity of moderate and severe HF from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in mean daily frequency of all HF from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in mean severity of all HF from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in the mean daily HF Score (frequency x severity) at Weeks 1, 2, 4, 8, 12 and 16
- Responder analyses: proportion of subjects with >50% and >80% reduction from baseline in mean daily HF frequency at Week 12
- Mean change from baseline in mean daily number of night-time awakenings (NTA) secondary to HF at Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in mean daily number of all NTA at Weeks 1, 2, 4, 8, 12 and 16
- Change from baseline in the Global and individual domain scores of the Pittsburgh Sleep Quality Index at Weeks 4, 8, 12 and 16
- Change from baseline in the Insomnia Severity Index score at Weeks 4, 8, 12 and 16
- Change from baseline in the HFRDIS scores at Weeks 2, 4, 8, 12 and 16
- Change from baseline in the MenQoL-I scores at Weeks 4, 8, 12 and 16
- Change from baseline in the Beck Depression Inventory II scores at Weeks 2, 4, 8, 12 and 16

### 2.5.2. Pharmacokinetic Variables

Exposure-response modelling will be undertaken on a number of efficacy and safety endpoints on an exploratory basis. These analyses will be described in a separate Exposure-Response Data Analysis Plan and the resulting data will be reported in a report separate from the CSR.

### 2.5.3. Safety Variables

Safety assessments performed during the study included physical examinations, measurement of vital signs, 12-lead electrocardiograms (ECGs), clinical laboratory evaluations including hematology, serum chemistry, and urinalysis, and monitoring of adverse events.

The safety endpoints include:

- Change from baseline in clinical laboratory assessments:
  - ✓ At Weeks 2, 4, 12 and 16 for haematology and urinalysis
  - ✓ At Weeks 2, 4, 8, 12 and 16 for clinical chemistry
- Change from baseline in vital signs (blood pressure, pulse rate, temperature) at Weeks 2, 4, 8, 12 and 16
- Change from baseline in weight, waist circumference and body mass index at Weeks 2, 4, 8, 12 and 16
- Proportion of subjects with clinically significant abnormal ECG findings at each visit
- Proportion of subjects with non-significant abnormal ECG findings at each visit
- Change from baseline at Weeks 2, 4, 8, 12 and 16 in ECG intervals (RR, PR, QT, QTc and QTcF)
- Proportion of subjects with absolute QTcF values by category at each visit: <450, >450 to <480, >480 to <500, >500 msec
- Proportion of subjects with change from baseline in ECG QTcF values by category at Weeks 2, 4, 8, 12 and 16: <0, >0 to <30, >30 to <60, >60 msec
- Change from baseline in the electronic Columbia Suicide Severity Rating Scale (eC-SSRS) at Weeks 4, 12 and 16
- Change from baseline to Weeks 12 and 16 in:
  - Serum concentration of bone-specific alkaline phosphatase (BALP)
  - Serum concentration of procollagen type 1 N-terminal propeptide (P1NP)
- Nature and severity of AEs
- Withdrawals due to an AE
- Use of concomitant medications

# 3. SUBJECT POPULATIONS

## 3.1. Population Definitions

The following subject populations will be evaluated and used for presentation and analysis of the data:

**Screen Analysis Set:** the screened population will include all subjects with informed consent.

**Randomized Analysis Set:** the randomized population will include all subjects with a randomization date and number.

**Safety Analysis Set**: All subjects who receive at least one dose of double-blind study drug irrespective of treatment received. Subjects will be analysed according to treatment received.

**Full Analysis Set (FAS):** All randomised subjects who received at least one dose of double-blind study drug, irrespective of treatment received, and have hot flush data for at least 7 days' worth of post-treatment assessments (i.e. requirement for the primary efficacy endpoint). Subjects will be analysed according to randomised treatment. This is the primary efficacy analysis set for the study.

**Modified Full Analysis Set (mFAS):** All subjects in the FAS excluding those who took non-hormonal treatments (anti-depressants, alpha-2 agonists, gabapentinoids and cannabis) or over the counter herbal treatments (intended to treat menopausal symptoms) that may have had a confounding effect on hot flash frequency and/or severity at any time during the study (defined as within 28 days before the start of Screening Visit 2 through to the Week 12 visit, inclusive). The medications entered in EDC have been reviewed by Sponsor's blinded Medical Expert, and the following final list of PREFERRED NAME was confirmed:
CLONIDINE
CLONIDINE HYDROCHLORIDE
CIMICIFUGA RACEMOSA
TRIFOLIUM PRATENSE
CIMICIFUGA RACEMOSA;HYPERICUM PERFORATUM
CANNABIS SATIVA
GABAPENTIN
PREGABALIN
AMITRIPTYLINE
NORTRIPTYLINE
CITALOPRAM
CITALOPRAM HYDROBROMIDE
ESCITALOPRAM
ESCITALOPRAM OXALATE
FLUOXETINE
FLUOXETINE HYDROCHLORIDE
PAROXETINE
PAROXETINE HYDROCHLORIDE
SERTRALINE
SERTRALINE HYDROCHLORIDE
BUPROPION
BUPROPION HYDROCHLORIDE
DESVENLAFAXINE
DESVENLAFAXINE SUCCINATE

DULOXETINE
DULOXETINE HYDROCHLORIDE
VENLAXAFINE
VENLAFAXINE HYDROCHLORIDE
MIRTAZAPINE
TRAZODONE
TRAZODONE HYDROCHLORIDE
VORTIOXETINE HYDROBROMIDE
CALCIUM;CIMICIFUGA RACEMOSA EXTRACT;GENISTEIN;HERBAL EXTRACT NOS;MAGNESIUM;PIPER METHYSTICUM RHIZOME
.. Subjects will be analysed according to randomised treatment. This is the sensitivity efficacy analysis set for the study planned only if more than 20% of subjects are excluded from the FAS.

**Per Protocol (PP) Set**: All subjects in the FAS who completed the 12 week treatment period excluding those identified as having relevant protocol deviations (see Section 3.2). The list of relevant protocol deviations leading to exclusion from the PP population has been finalized during the blinded data review prior to database lock and unblinding of treatment allocation. Subjects will be analysed according to randomised treatment. This is the secondary efficacy analysis set for the study.

## 3.2. Protocol Deviations

Protocol deviations are defined as a deviation from the approved protocol.

Prior to database lock, NeRRe will be responsible for producing the final protocol violation file (formatted as a Microsoft Excel file), in collaboration with the data monitoring group; this file will include a description of the protocol violation and the categorization as minor/major, relevant / not relevant. Deviations classified as relevant will result in exclusion of the subject from the Per Protocol Population. This file will be finalized and signed prior to hard database lock, after the blinded data review meeting (BDRM).

In addition, the following protocol deviations will be programmatically derived by Cytel:

1) Category = Non-compliance with diary completion

   - PD Reference 3.9: Less than 70% compliance with hot flush diary completion at Baseline (Visit 3).

   - PD Reference 3.10: Less than 70% compliance with hot flush diary completion at Week 4 (Visit 5).

   - PD Reference 3.11: Less than 70% compliance with hot flush diary completion at Week 12 (Visit 7).

   - PD Reference 3.12: Less than 70% compliance with hot flush diary completion at Week 2 (Visit 4), Week 8 (Visit 6) or Week 16 (Visit 8).

   Compliance will be calculated for Baseline, Week 2, Week 4, Week 8, Week 12 and Week 16, using the same 7 days (or less than 7 days if this is the case per the rules defined in section 4.3.2) used in the analysis of HF frequency & severity.

The 70% cut-off essentially means that at least 10 of the 14 possible diary entries will need to have been completed for the subject to be in the PP population.

2) Category = Non-compliance with study medication

PD Reference 4.2: Subject <80% compliant with once-daily dosing of IMP.

IMP compliance will be based on the number of capsules taken (section 4.8.1). The compliance calculation should be across the whole 12-week double-blind treatment period (not visit specific). IMP should not be missing.

Relevant protocol deviations will be summarised by treatment group for the Randomised Analysis Set. All protocol violations will be presented in the data listings.

# 4. STATISTICAL METHODS

## 4.1. Sample Size Justification

The total sample size for the trial yields high power for the expected large difference from placebo in hot flush reduction from baseline and adequate power for moderate difference.

Assuming a common standard deviation (SD) of 4.4, N=27 per treatment group has ~95% power (alpha=0.05 2-sided) via trend test across all doses including placebo if the true underlying dose response is non-decreasing from a reduction of 4 moderate or severe hot flushes per day on placebo up to 8 on the highest dose of NT-814. N=33 per treatment group is required for 95% power for pairwise comparison of the highest dose to placebo under these same assumptions. Hence, a total of 150 completed subjects (30 x 5) was used for the simulations of adaptive design. [Note, the actual performance / power for the adaptive designs may be higher than the values cited in this report since additional subjects will be added to selected dose groups to better inform on safety.] The actual sample size is determined by the adaptive dose-finding design algorithm (See Appendix B from the protocol). In addition, the actual sample size is increased by 10% to allow for subjects who withdraw prematurely and so the initial sample size is 165 subjects.

The interim reviews will also include a review of the estimates used in defining the starting sample size. If the observed variance or delta are lower than the original estimates the sample size for the dose groups that are continuing to be evaluated may be increased. For this reason, the protocol permits a total of up to 300 subjects to be randomised into the study.

## 4.2. General Statistical Methods and Data Handling

### 4.2.1. General Methods

Tabulations will be produced for appropriate demographic, baseline, efficacy and safety parameters. For categorical variables, summary tabulations of the number and percentage within each category of the parameter will be presented. Percentages will be calculated based on the number of non-missing values, except if indicated differently. In case of counts equal to zero, reporting will be presented as 0 (with no percentage). For continuous variables, the mean, median, standard deviation, minimum and maximum values will be presented.

For specific categorical efficacy parameters, two-sided 95% confidence interval might be presented.

Formal statistical hypothesis testing will be performed on the primary and secondary efficacy endpoints with all tests conducted at the 2-sided, 0.05 level of significance. Summary statistics will be presented, as well as confidence intervals on selected parameters, as described in the sections below. P-values will be reported with 4 decimals places.

### 4.2.2. Computing Environment

All descriptive statistical analyses will be performed using SAS statistical software (Version 9.4), unless otherwise noted. Medical History and adverse events will be coding using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary Version 21.1 or in the latest version available prior to DBL. Concomitant medications will be coded using World Health Organization (WHO) Drug Version Sep2018 B3 Global or the latest version available prior to DBL.

4.2.3. Adjustments for Covariates

All efficacy endpoints will be analysed by statistical models including treatment, score/endpoint at baseline, randomisation stratification factor (region: North America / Europe) as fixed effect covariates.

4.2.4. Multiple Comparisons/Multiplicity

No adjustment for multiple comparisons will be used for this Phase 2 study.

4.2.5. Subpopulations

The descriptive analysis of the co-primary endpoints analysis will be repeated by region (North America or Europe).

4.2.6. Withdrawals, Dropouts, Loss to Follow-up

Subjects who withdrew from the study after randomisation will not be replaced.

4.2.7. Missing, Unused, and Spurious Data

In general, there will be no substitutions made to accommodate missing data points. All data recorded on the CRF will be included in data listings that will accompany the clinical study report.

Sensitivity analyses may be performed, and are detailed in the corresponding efficacy endpoint sections.

Clinical laboratory values as "<X" where X is the numerical value of the limit of quantification (LOQ) established by the laboratory will be imputed by LOQ/2 for summaries in the tables.

Clinical laboratory values as ">X" where X is the numerical value of the limit of quantification (LOQ) established by the laboratory will be imputed by LOQ for summaries in the tables.

4.2.8. Study Day

Study Day is the number of days since the date of the administration of the first dose of randomised study treatment (Study Day 1). If the assessment date is after the date of the first dose, the study day is calculated as (date of assessment - date of the first randomised treatment intake date+ 1). If the assessment date is prior to the date of the first dose, the study day is calculated as (date of assessment - date of the first randomised treatment intake date). All assessments prior to Study Day 1, including the Screening visits 1 and 2, will have negative study days (i.e. no Day 0).

4.2.9. Date of last assessment

Date of last assessment will be the date of the last visit per subject as recorded on the subject visit date CRF (SV (Study Visits) SDTM dataset).

4.2.10. Planned and actual treatment

Planned treatment will be the one assigned per the randomization list. Possible treatment groups are:

- Placebo
- NT-814 40 mg once a day

- NT-814 80 mg once a day
- NT-814 120 mg once a day
- NT-814 160 mg once a day

Actual (received) treatment might be different if the kit(s) actually dispensed does not correspond to the kit(s) allocated by IWRS. Thus, actual treatment groups will be derived – for patients belonging to the Safety Analysis Set – according to the majority of treatment dispensed to the patient. Indeed, kits are dispensed at Baseline, Week 4 and Week 8 visits. In case of there is no treatment given in majority, then the actual treatment group will correspond to the kit actually dispensed to the subject at Visit 3 for the first treatment period from baseline to Week 4. If the information on the kit number is not available at the time of the analysis, the actual treatment will be considered as equal to the planned one.

In practice, this means that a dispensing error could result in a subject being analysed in an actual treatment group different from its planned treatment group in case of dispensing error.

Any occurrence of incorrectly administered treatment should be recorded as a PD, and will be reviewed at the BDRM to assess on an individual case by case basis which treatment group the subject will be allocated to.

#### 4.2.11. Baseline definition

Baseline is defined as the data most recently collected prior to the first randomised treatment intake date/time. If time of the corresponding data (e.g. Questionnaires) is not collected then, we will assume that the assessment at Visit 3 is performed as planned, i.e. prior to the treatment intake.

The baseline assessment for hot flushes will be calculated using the last 7 days (not necessarily consecutive days) with an available data in the evening and/or the morning of the baseline diary completion period. A diary day is comprised of the evening entry of this day and the morning entry of the following day, in that order.

Mean daily frequency = Sum of number of hot flushes filled in the diary during the last 7 diary days (with at least one available data in the evening and/or morning) divided by 7.

Mean weekly severity = (number of moderate hot flushes for 7 days) x 2 + (number of severe hot flushes for 7 days) x 3] / (total number of moderate to severe hot flushes over 7 days).

Mean daily HF score = Sum of (frequency x severity) filled in the diary during the last 7 days (with at least one available data in the evening and/or morning) divided by 7. Severity is graded by the women from 1 to 3 (1 = mild; 2 = moderate; 3 = severe).

For the primary efficacy endpoint analyses and several of the secondary endpoint analyses, only moderate and severe HF data are included in the calculations. For these analyses the mild data recorded in the diary will be excluded from the baseline values.

A diary day comprises an evening entry and a morning entry, in that order. If a diary has been completed on a particular day but is missing data on HF that day, then it will be assumed that there was no HF on that day. This diary will not be used for the 7 days of baseline data. Only diary with available data will be used in baseline derivation.

#### 4.2.12. Visit Windows

Visit windows are defined in Table 2, and are applicable for efficacy (except diary data) and safety.

**Table 2 Evaluation Intervals for Efficacy and Safety Analysis**

| Evaluation | Targeted time point | Protocol-Specified Interval | Interval for Analysis |
|---|---|---|---|
| Screening Visit 1 | Day -21 | Day -21 | Day -18 or earlier |
| Screening Visit 2* | Day -14 | Day -16 to Day -12 | Day -17 to Day -4 |
| Baseline | Day 1 | Day 1 | Day –3 to Day 1 |
| Week 2 | Day 15 | Day 12 to Day 18 | Day 2 to Day 22 |
| Week 4 | Day 29 | Day 26 to Day 32 | Day 23 to Day 43 |
| Week 8 | Day 57 | Day 53 to Day 61 | Day 44 to Day 71 |
| Week 12 | Day 85 | Day 81 to Day 89 | Day 72 to Day 99 |
| Week 16 FU | Day 113 | Day 108 to Day 118 | Day 100 or later |

* At least 6 days after Screening Visit 1 to allow at least 6 days of paper diary data to be completed.
Study day is calculated per section 4.2.8.

In case more than one measurement per time period is captured, the closest to the targeted time point will be considered for analysis. If two measurements occur at the same distance from the target day for a particular analysis window then the first measurement will be considered for analysis. Unscheduled visits will be taken into account. In case of withdrawal, all data from the early termination visit will be assigned under its corresponding analysis visit according to its day.

Actual dates and times will be used for pharmacokinetic summaries rather than nominal days and times.

Any visits occurring outside visit windows defined in the protocol should be recorded as a PD and will be reviewed at the BDRM to assess on an individual case by case basis to which visit the data should be allocated.

## 4.3. Interim Analyses

The first IA will be conducted after efficacy data from frequency and severity of HF is collected through Week 4 from 30 patients randomized in equal proportions to the four NT-814 doses and placebo.

Based on DRC review of the IA results, randomization ratios for subsequent enrolled patients may be changed to optimize the allocation of subjects to doses with target levels of weekly average daily HF frequency reduction from baseline of 6 and 8; and HF severity reduction from baseline of 0.66 and 0.88 (Study RELENT-1). Adaptive design algorithm will be based on Bayesian Emax dose-response modelling [1] and / or T-statistic adaptive dose-finding design [2]. The DRC will have dose assignment recommendations by each of those adaptive dose-finding algorithms and assign the randomization ratio that best integrates the objectives of the trial and the adaptive design results. Randomisation ratio recommendations will also take into consideration emerging safety findings as well as the need to allocate sufficient subjects to a dose of interest to enable an adequate assessment of safety to be made.

Subsequent IA's will be conducted similarly, after Week 4 data from each cohort of ~30 additional patients become available.

The DRC will evaluate stopping criteria per the IA algorithms as described in Appendix B of the protocol and decide on stopping assignment of subjects to specific doses or stopping the trial for an efficacy conclusion and dose choice for subsequent study or futility.

The review of efficacy will be based only on the frequency and severity of HF recorded by subjects in the electronic diary (eDiary).

The review of safety will be based on adverse event and early withdrawal information.

For each IA, a cut-off date will be applied such that data will be included until the date the last patient in the analysed cohort completes the Week 4 visit.

4.3.1. Demographic and baseline data

Demographic data and menopause characteristics will be summarised by treatment group for the Safety Analysis Set (see sections 4.5.1 and 4.5.2 for the details).

4.3.2. Efficacy analysis

At each interim review, summaries by dose group will be provided on the FAS for each of the following:

- Mean daily frequency of moderate and severe HF at baseline, mean daily frequency of moderate and severe HF at each study week, including Week 4 and mean change and percent change from baseline in the mean daily frequency of moderate and severe HF at each study week.
- Mean severity of moderate and severe HF at baseline, mean severity of moderate and severe HF at each study week, including Week 4, and mean change from baseline in the severity of moderate and severe HF at each study week.

If available, summaries of the same data at later time-points (Week 8 and Week 12) will also be reviewed.

Mean daily frequency at each study visit will be calculated using the same method as the one described in section 4.2.11 based on the last 7 days with at least one available data in the evening and/or morning before the corresponding visit.

Mean weekly severity at each study visit will be calculated using the following:

[(number of mild hot flushes for 7 days) x 1 + (number of moderate hot flushes for 7 days) x 2 + (number of severe hot flushes for 7 days) x 3] / total number of mild, moderate and severe hot flushes over 7 days.

If the number of hot flushes is equal to 0, then the mean severity will be set to 0.

These 7 last days with available data should be within the 14 days prior to the CRF study visit, so in the following window: [CRF Visit date – 14; CRF Visit date], with at least one day within

the last 7 days prior to the study visit. In case of withdrawal, the day this ET visit happened will be checked to see in which windows it is and then to assign the corresponding timepoint. If a CRF visit has already been done at that timepoint, then the CRF visit date will be used. If less than 7 days fulfil these rules, only the correct data will be used in the derivation, and formula will be adjusted using the right number of days as denominator. Morning diary data on the day of a Visit will be included in the derivation as the morning diary is actually associated with the previous calendar day (a complete day is an evening then a morning diary).

In addition, emerging data will be evaluated via the two adaptive following algorithms to evaluate the emerging data at Week 4:

(i) a Bayesian adaptive Emax (S-shaped) dose-response model fit, and
(ii) a model based on isotonic regression (T-stat)

The models have been developed using the following prior assumptions:

1) <u>On frequency of hot flushes:</u>

- The mean reduction in the mean daily number of moderate and severe HFs in the placebo group is 4 per day
- Two target levels of test treatment: mean reduction from baseline in the mean daily number of moderate and severe HFs of 6 and 8
- Common standard deviation (SD) of 4.4
- Stopping guidelines for Emax if (a) the probability is > 90% that the mean hot flush reduction is at least 8 at that dose, or (b) the probability is <20% that the mean hot flush reduction is at least 5 (i.e., >80% probability that the mean hot flush reduction is <5).

2) <u>On severity of hot flushes:</u>

- The mean reduction in the severity of moderate and severe HFs in the placebo group is 0.44 per day
- Two target levels of test treatment: mean reduction from baseline in the severity of moderate and severe HFs of 0.66 and 0.88
- Common standard deviation (SD) of 0.55
- Stopping guidelines for Emax if (a) the probability is > 90% that the mean severity hot flush reduction is at least 0.88 at that dose, or (b) the probability is <20% that the mean severity hot flush reduction is at least 0.22 (i.e., >80% probability that the mean hot flush reduction is <0.22).

Emerging data (mean change from baseline in the frequency/severity of moderate and severe HF at Week 4) will be entered in COMPASS software using the Execute module. The following steps will be followed:

- Raw data from eDiaries will be provided to the Cytel blinded team by ERT, the CRF data will be provided by the Pharm Olam Data Management team
- Efficacy endpoints to be presented (see above) will be generated by the Cytel blinded team using the derivation described in this SAP
- Statistical descriptive tables for these endpoints will be produced by the Cytel blinded team using a dummy randomisation list

- Programs will be rerun by the Cytel unblinded team using the real randomisation list provided to the Cytel unblinded statistician
- Endpoint data for each patient along with their treatment arm will be entered into COMPASS.

Prior distributions, defined above and used for the simulation, will also be entered in the software. COMPASS has been developed as a software package for design of adaptive early stage trials within Cytel's suite of software programs for design, analysis, and implementation of clinical trials.

Outputs from the software after having the randomisation code applied will be provided to the DRC members. The outputs comprise:

- Summary tables showing: subject allocation, observed and model-estimated mean results, by dose and pooled SD's, posterior probability (for Emax) or conditional power (for T-stat) and summary statistics
- New subject data with a new recommended randomisation ratio, considered the most efficient per the method.

A summary of the results provided by COMPASS included the allocation probability (related to the next randomisation ratio) recommended from the Emax and T-stat analyses, based on the frequency and severity of HF, will be provided by the unblinded statistician to the DRC members in a separate report. The DRC will take into consideration these results in recommending the randomisation ratio for the next cohort of subjects.

#### 4.3.3. Safety analysis

An overview of adverse events will be presented and will include the number and percentage of patients in the different categories indicated in section 4.8.2.

The following information will be listed by treatment group for review at each meeting:

- Serious adverse events
- Withdrawals due to adverse events
- Other early study withdrawals with reason
- Severe adverse events
- Adverse events

Adverse event summaries will include all available data and will not be limited just to data through Week-4.

## 4.4. Subject Disposition

A tabulation of subject disposition will be provided on the Screen Analysis Set and will include:

- the number of patients screened,
- the number of screen failures,
- the number of patients randomized,

- the number of patients who received randomised treatment.

A tabulation of randomised treatment and study completion/discontinuation status will be provided on the Randomised Analysis Set and will include

- the number of patients randomized,
- the number of patients who received randomised treatment.
- the number of patients who completed the treatment
- the number of patients who withdrew prior to completing the treatment, and reasons for withdrawal,
- the number of patients who completed the study,
- the number of patients who withdrew prior to completing the study, and reasons for withdrawal.

An overview of the number of patients included in each population together with reason for screen failure or exclusion from populations will be produced for the Randomised Analysis Set.

A by-subject listing of study and treatment completion information, including the reason for premature study withdrawal, if applicable, will be presented. All study and visit dates will be also listed.

The number of patients by country, and by site within each country will also be summarized on the Safety Analysis Set.

## 4.5. Demographic and Baseline Characteristics

### 4.5.1. Demographics characteristics

Demographics information will be summarized by treatment group using descriptive statistics for the Safety Analysis Set and will include:

- Age at screening (years)
- Ethnicity
- Race
- Region
- Weight [kg]
- Height [cm]
- Body Mass Index (BMI) [kg/m2]: defined as Weight (kg)/(Height (m))^2

Weight collected at baseline (Visit 3) and height collected at Visit 2 will be used.

The following conversions will be applied:

Weight (kg) = 0,45359237 * Weight (lb)

Height (cm) = 2,54 * Height (Inches)

Demographic data will also be provided in data listings.

### 4.5.2. Baseline Disease characteristics

Baseline Disease characteristics will be summarized by treatment group using descriptive statistics for the Safety Analysis Set and will include:

- Duration of menopause (years) as continuous summary and in categories [<5 years; ≥5 to <10 years; ≥10 years].
- Age at menopause Onset (years) as continuous summary and in categories [<50 years; ≥50 years].
- Menopause history

Computation of Duration of menopause and Age at menopause onset

The CRF requests to record date of the last menstrual period. However, a partial date is accepted.

Imputation will be performed in case the date of the last menstrual period is partial:

- If only Day is missing, it will be imputed as 1st of the month
- If day & month are missing, it will be imputed as 1st January
- If year is missing, no imputation will be done.

After imputation, the date of the last menstrual period must be more than 6 weeks prior to the date of Informed Consent (IC).

Duration of menopause (y) = (Informed Consent date – Last menstrual period date) +1/365.25

As year of birth is not collected, it will be estimated using:

Year of birth = Informed Consent year – Age on day of screening

Age of menopause (y) = Last menstrual period year – Year of Birth

Baseline data will also be provided in data listings.

### 4.5.3. Medical history

The medical history conditions will be summarized by treatment group for the Safety Analysis Set as frequencies and percentages according to the System Organ Class (SOC) and Preferred Term (PT) levels.

Medical history will be sorted by decreasing order of frequency by SOC and PT of the Total group. The listing will display the SOC, PT, and the verbatim text from the study Investigators.

## 4.6. Efficacy Evaluation

Efficacy analyses will be conducted using the FAS population. The co-primary efficacy endpoints will, additionally, be analysed on the PP analysis set and on the mFAS analysis set (if at least 20% of subjects from the FAS used relevant treatments within the 28 days before screening). The statistical methods will focus on summarizing the data collected by visit using appropriate tabular and graphical presentations. Diary data, individual items and scores for questionnaires will be displayed in listings. Plots will be produced for some endpoints, as detailed in the following sections.

### 4.6.1. Co-primary analysis

The co-primary endpoints are the change from baseline in the mean daily frequency and mean weekly severity of moderate and severe hot flushes at Weeks 4 and 12.

For each timepoint (Baseline, Week 1, Week 2, Week 4, Week 8, Week 12) summaries will be produced which use data from the 7 days with at least one available eDiary data entry in the evening and/or the morning before the Visit date:

- These 7 days do not need to be consecutive.
- As the data entered in the morning diary is assigned to the previous calendar day (a diary day comprises an evening and a morning entry, in that order) the morning diary with the same calendar date as the Visit day may be included in the calculation.
- For Week 1, the first 7 post-baseline days with at least one available data in the evening and/or the morning will be used.

The methods for calculating the endpoint are the same as described in section 4.2.11.

Absolute and changes from baseline in the mean daily frequencies and mean weekly severity of moderate and severe hot flushes will be summarised by treatment group. Percent changes from baseline in the mean daily frequencies will be also summarised.

The change from baseline will be calculated as follow:

Mean daily frequency (or mean severity) at week 4 (or Week 12) – Mean daily frequency (or mean severity) at baseline.

The percent change from baseline will be calculated as follow:

(Change from baseline / Mean daily frequency at baseline) * 100.

The mean change from baseline and the corresponding 95% confidence interval (CI) will be displayed graphically by visit on the observed values and by treatment groups.

Bar graphs of the change from baseline at Week 4 and Week 12 will be produced for each treatment group.

The change from baseline endpoint will be analysed using a Mixed-Effect Model Repeated Measures (MMRM) incorporating post-randomization data collected up to treatment discontinuation at weeks 12 (Week 1, 2, 4, 8 and 12 will be included; Week 16 visit data will not be included in the model) and with consideration of the variance-covariance matrix of the repeated measures.

This method allows for a general unstructured variance-covariance matrix and will include data from subjects with incomplete data from some scheduled time points. The model will be implemented in SAS using the MIXED procedure and will include the change from baseline as the dependent variable. The fixed effects in the model will include independent variables of randomized treatment, visit (nominal post-baseline visits as per the schedule of assessments) and treatment-by-visit interaction, along with the following baseline covariates: Baseline frequency or severity, Region (section 4.2.3). Visits will be treated as a repeated variable within a patient. Patient, treatment and visit will be treated as factor variables. An unstructured variance-covariance structure will be applied to model the within-patient errors. The model will be fitted using the Restricted Maximum Likelihood (REML) method. Denominator degrees of

freedom will be estimated using Kenward-Roger's approximation.

Pairwise statistical comparisons are planned for each NT-814 dose group (40 mg, 80 mg, 120 mg and 160 mg) versus placebo. To estimate the difference between the treatment groups in mean change from baseline to Week 4 and Week 12, a treatment-by-visit interaction contrast will be constructed (i.e., the treatment group contrast at Week 4 and 12). On the basis on this analysis, least square means, standard errors, and the 95% confidence interval for the treatment difference (at Week 4 / 12, primary endpoint timepoints) will be reported.

SAS code:

```
PROC MIXED DATA= dataset;
CLASS region avisitn treat subjid;
MODEL chg=base treat region avisitn treat*avisitn / ddfm=KENWARDROGER;
REPEATED avisitn / subject=subjid type=un;
LSMEANS treat*avisitn / cl pdiff;
ODS OUTPUT lsmeans=lsmeans_ tests3 = tests3_ diffs = diffs_
convergenceStatus=convergenceStatus_;
WHERE avisitn > xx;
RUN;
* chg represents the mean change from baseline variable;
* base represents the mean daily frequency (or severity) at baseline;
* treat represents the treatment group;
* avisitn represents the analysis visits as defined in section 4.2.12. Only the
post-baseline visits must be included
* region represents the categorical covariate related to stratification factors;
Further options to control the output may be added.
From ODS OUTPUT, using the correct visit/treatment selection, the outputs will
allow getting the following:
- lsmeans_: LSM, SE and 95% CI for each treatment at each visit
- diffs_: Difference in LSM, 95% CI and p value
- tests3_: overall p values for the fixed effects.
```

The assumption of normality (to ensure that the parametric models are appropriate) will be explored via the visual checks that are based on a normal probability plot of the residuals against their values expected on the normal distribution. If the data are normally distributed then the residuals will roughly form a straight line on the normal plot. If the plotted data deviates markedly from a straight line then it is likely that the data are not normally distributed. In that case, a non-parametric analysis will be used as a sensitivity analysis to complement the parametric MMRM results. A Wilcoxon rank-sum test will be performed at each timepoint (Week 4 and 12) and for each pairwise treatment comparison.

SAS code:

```
PROC NPAR1WAY DATA= dataset WILCOXON;
CLASS treat;
VAR chg;
EXACT Wilcoxon;
WHERE avisit = "Week 4" and treat in ("Placebo","NT-814 40mg");
RUN;
* chg represents the mean change from baseline variable;
* treat represents the treatment group;
* avisit represents the analysis visits as defined in section 4.2.12.
```

In addition, as a sensitivity analysis, an analysis of covariance (ANCOVA) including terms for treatment group, along with the following baseline covariates: Baseline frequency or severity, Region (section 4.2.3). This analysis will be performed at each timepoint (Week 4 and Week 12).

SAS code:

```
PROC MIXED DATA=dataset;
CLASS treat region;
MODEL chg = treat base region;
LSMEANS treat / pdiff (ref='Placebo') cl;
ODS OUTPUT lsmeans=lsmeans_ tests3 = tests3_ diffs = diffs_
convergenceStatus=convergenceStatus_;
WHERE avisit = "Week 4";
RUN;
* base represents the baseline value of the endpoint;
* chg represents the change from baseline variable;
* treat represents the treatment group;
* region represents the categorical covariate related to stratification factors.
Further options to control the output may be added. See MMRM notes for ODS
OUTPUT.
```

As an ad-hoc analysis, treatment mean comparisons to placebo may be made via estimated means from a 4-parameter Emax model fit and trend test for increasing response with increasing dose based on isotonic regression modelling at Week 4 and Week 12 using the same methodology as described in section 4.3.2.

4.6.2. Secondary analysis

4.6.2.1. Other hot flush frequency and severity secondary endpoints

The following hot flush frequency and severity secondary endpoints will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

- Mean change from baseline in mean daily frequency of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in mean weekly severity of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in mean daily frequency of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in mean weekly severity of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in the mean daily Hot Flush Score (frequency x severity) at Weeks 1, 2, 4, 8, 12 and 16

The above endpoints will be derived using the 7 last days with at least one available data in the evening and/or the morning before the corresponding visit using the same method described in

section 4.2.11. For Week 1, the first 7 post-baseline days with at least one available data in the evening and/or the morning will be used.

The mean change from baseline and the corresponding 95% CI will be displayed graphically by visit on the observed values and by treatment groups.

The proportion of responders, defined as a reduction from baseline of >=50% and >=80% on the average mean daily frequency of moderate and severe hot flushes by week will be summarised at each evaluation time.

The proportion of responders will be calculated using the percent change from baseline as below at a visit Week x:

> Percent change = (change from baseline in mean daily frequency of moderate and severe HF from baseline to Week x / Mean daily frequency of moderate and severe HF at baseline) * 100

A subject will be considered as responder with reduction of >=50% (or >=80%) if the percent change is <=-50 (or <=-80).

The Week 4 and Week 12 response will be analysed using a logistic regression. The model will be implemented in SAS using the LOGISTIC procedure and will include the responder variables as response variable. The fixed effects in the model will include independent variable of randomized treatment, along with the following baseline covariates: Baseline frequency of moderate and severe HF, Region.

The odds ratio will be used as a measure of association between treatment and response and will be calculated such that an odds ratio >1 is favourable for NT-814. The 95% confidence interval for the odds ratio assumes asymptotic normality of the Wald estimate for the regression coefficient. The Wald p-value associated to the treatment covariate in the logistic regression will be provided. In addition, as a sensitivity analysis, the adjusted relative risk of response will be calculated from a Cochran-Mantel-Haenszel test adjusted on the categorized randomization stratification factor: Region (North America/Europe).

SAS code:

```
PROC LOGISTIC DATA= dataset;
CLASS treat (ref='Placebo') region / param=ref;
MODEL response (event='1') = base treat region;
ODS OUTPUT OddsRatios=OddsRatios_ ParameterEstimates=ParameterEstimates_;
RUN;
* base represents the baseline value of the endpoint;
* response represents the response variable;
* treat represents the treatment group;
* region represents the categorical covariate related to stratification factors;
Further options to control the output may be added.
From ODS OUTPUT, using the correct selection, the outputs will allow getting the
following:
- OddsRatios_: Odd ratio and 95% CI for each treatment
- ParameterEstimates_: p value.


PROC FREQ DATA = dataset;
Tables region*treat*response / cmh;
```

```
WHERE treat in ("Placebo","NT-814 40mg");
RUN;
* The p-value would be for "Row Mean Scores Differ" in the SAS PROC FREQ output.
```

The following night time awakenings secondary endpoints will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

- Mean change from baseline in mean daily number of night time awakenings secondary to hot flush at Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in mean daily number of all night time awakenings at Weeks 1, 2, 4, 8, 12 and 16

The above endpoints will be derived using the 7 last days with at least one available data in the morning before the corresponding visit using the same method than the one described in section 4.2.8.

Night time awakenings secondary to hot flush correspond to severe hot flash recorded on the morning diary, and all night time awakenings correspond to the data recorded in "Total number of times you woke up last night?" field from eDiary recorded in the morning. Number of NTAs secondary to HF can't be higher than number of all NTAs.

The mean change from baseline and the corresponding 95% CI will be displayed graphically by visit on the observed values and by treatment groups.

#### 4.6.2.2. Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) is a self-rated questionnaire which assesses sleep quality and disturbances over a 1-month time interval. Nineteen individual items generate seven "component" scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction, each scored 0 (no difficulty) to 3 (severe difficulty). The sum of scores for these seven components yields one global score (range 0 to 21). Higher scores indicate worse sleep quality.

Scoring:

| | |
|---|---|
| **PSQIDURAT** | **DURATION OF SLEEP**<br>IF Q4 ≥ 7, THEN set value to 0<br>IF Q4 < 7 and ≥ 6, THEN set value to 1<br>IF Q4 < 6 and ≥ 5, THEN set value to 2<br>IF Q4 < 5, THEN set value to 3<br>Minimum Score = 0 (better); Maximum Score = 3 (worse) |

**PSQIDISTB**

**SLEEP DISTURBANCE**
IF Q5b + Q5c + Q5d + Q5e + Q5f + Q5g + Q5h + Q5i + Q5j (IF Q5JCOM is null or Q5j is null, set the value of Q5j to 0) = 0, THEN set value to 0

IF Q5b + Q5c + Q5d + Q5e + Q5f + Q5g + Q5h + Q5i + Q5j (IF Q5JCOM is null or Q5j is null, set the value of Q5j to 0) ≥ 1 and ≤ 9, THEN set value to 1

IF Q5b + Q5c + Q5d + Q5e + Q5f + Q5g + Q5h + Q5i + Q5j (IF Q5JCOM is null or Q5j is null, set the value of Q5j to 0) > 9 and ≤ 18, THEN set value to 2

IF Q5b + Q5c + Q5d + Q5e + Q5f + Q5g + Q5h + Q5i + Q5j (IF Q5JCOM is null or Q5j is null, set the value of Q5j to 0) > 18, THEN set value to 3

Minimum Score = 0 (better); Maximum Score = 3 (worse)

**PSQILATEN**

**SLEEP LATENCY**
**First, recode Q2 into Q2new thusly:**
IF Q2 ≥ 0 and ≤ 15, THEN set value of Q2new to 0
IF Q2 > 15 and ≤ 30, THEN set value of Q2new to 1
IF Q2 > 30 and ≤ 60, THEN set value of Q2new to 2
IF Q2 > 60, THEN set value of Q2new to 3

**Next**
IF Q5a + Q2new = 0, THEN set value to 0
IF Q5a + Q2new ≥ 1 and ≤ 2, THEN set value to 1
IF Q5a + Q2new ≥ 3 and ≤ 4, THEN set value to 2
IF Q5a + Q2new ≥ 5 and ≤ 6, THEN set value to 3

Minimum Score = 0 (better); Maximum Score = 3 (worse)

**PSQIDAYDYS**

**DAY DYSFUNCTION DUE TO SLEEPINESS**
IF Q8 + Q9 = 0, THEN set value to 0
IF Q8 + Q9 ≥ 1 and ≤ 2, THEN set value to 1
IF Q8 + Q9 ≥ 3 and ≤ 4, THEN set value to 2
IF Q8 + Q9 ≥ 5 and ≤ 6, THEN set value to 3
Minimum Score = 0 (better); Maximum Score = 3 (worse)

**PSQIHSE**

**SLEEP EFFICIENCY**
Diffsec = Difference in seconds between day and time of day Q1 and day Q3
Diffhour = Absolute value of diffsec / 3600
newtib =IF diffhour > 24, then newtib = diffhour – 24
IF diffhour ≤ 24, THEN newtib = diffhour
(NOTE, THE ABOVE JUST CALCULATES THE HOURS BETWEEN GNT (Q1) AND GMT (Q3))
tmphse = (Q4 / newtib) * 100

IF tmphse ≥ 85, THEN set value to 0
IF tmphse < 85 and ≥ 75, THEN set value to 1
IF tmphse < 75 and ≥ 65, THEN set value to 2
IF tmphse < 65, THEN set value to 3
Minimum Score = 0 (better); Maximum Score = 3 (worse)

| | |
|---|---|
| **PSQISLPQUAL** | **OVERALL SLEEP QUALITY**<br>Q6<br>Minimum Score = 0 (better); Maximum Score = 3 (worse) |
| **PSQIMEDS** | **NEED MEDS TO SLEEP**<br>Q7<br>Minimum Score = 0 (better); Maximum Score = 3 (worse) |
| **PSQI** | **TOTAL**<br>DURAT + DISTB + LATEN + DAYDYS + HSE + SLPQUAL + MEDS<br>Minimum Score = 0 (better); Maximum Score = 21 (worse)<br>Interpretation: TOTAL ≤ 5 associated with good sleep quality<br>TOTAL > 5 associated with poor sleep quality |

Analysis:

For the 7 "component" scores and the global score, absolute and changes from baseline will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

4.6.2.3. Insomnia Severity Index

The Insomnia Severity Index (ISI) is a brief self-report questionnaire assessing the nature, severity, and impact of insomnia. The ISI comprises 7 items assessing the perceived severity of difficulties initiating sleep, staying asleep, and early morning awakenings, satisfaction with current sleep pattern, interference with daily functioning, noticeability of impairment attributed to the sleep problem, and degree of distress or concern caused by the sleep problem. Subjects rate each item on a scale of 0 to 4, yielding a total score ranging from 0 to 28.

Scoring:

The total score is calculated by adding the scores for all seven items. Higher scores indicate severe insomnia.

Analysis:

Absolute and changes from baseline will be summarised by treatment group and scheduled visit (as applicable) for the total score. In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

4.6.2.4. Hot Flash Related Daily Interference Scale (HFRDIS)

The HFRDIS is a 10-item, self-report questionnaire assessing the impact of hot flushes on a woman's life during the past week. For each of the 10 items, subjects rate how much hot flushes have interfered with that aspect of their life on a scale of 0 (not at all) to 10 (very much so).

Scoring:

The total score is calculated by adding the scores for all ten items. Higher scores indicate greater interference.

Analysis:

For each of the 10 items and the total score, absolute and changes from baseline will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

4.6.2.5. MenQoL-I (1 month recall version)

The MenQoL-I (Menopause-specific Quality-of-Life Questionnaire Intervention Version) is a validated questionnaire used to measure condition-specific quality of life in menopausal women. It is composed of 32 items across 4 domains (physical, vasomotor, psychosocial and sexual). For each item, subjects record whether they have experienced the problem in the past month, and if so, they rate how bothered they were by the problem on a scale of 0 (not at all bothered) to 6 (extremely bothered). The item responses can then be converted into analysis scores and an overall questionnaire score.

Scoring:

The scale contains four domains:

- Vasomotor: items 1 to 3
- Psychosocial: items 4 to 10
- Physical: items 11 to 26, 30, 31, 32
- Sexual: items 27 to 29

Convert the item scores to a score ranging from 1 to 8 in the following manner:

| Subject response | Analysis score |
|---|---|
| No | 1 |
| 0 | 2 |
| 1 | 3 |
| 2 | 4 |
| 3 | 5 |
| 4 | 6 |
| 5 | 7 |
| 6 | 8 |

The score for each domain is calculated by the mean of the items contained in each (ranging from 1 to 8).

The overall questionnaire score is the mean of the domain scores.

Analysis:

For the 4 domain scores and the overall score, absolute and changes from baseline will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

4.6.2.6. Beck Depression Inventory II

The Beck Depression Inventory II (BDI-II) is a 21-item questionnaire assessing the intensity of depressive symptoms over the past 2 weeks. It is composed of items relating to symptoms of depression such as hopelessness and irritability, cognitions such as guilt or feelings of being punished, as well as physical symptoms such as fatigue, weight loss, and lack of interest in sex.

Scoring:

Subjects rate each item on a scale of 0 to 3, with the total score ranging from 0 to 63, with a higher score suggesting more severe depressive symptoms.

Analysis:

For the total score, absolute and changes from baseline will be summarised by treatment group and scheduled visit (as applicable). In addition, the same approach as for the primary endpoint (MMRM incorporating post-randomization data collected up to treatment discontinuation at week 12) will be used to provide estimates of the change from baseline to weeks 4 and 12.

4.6.2.7. Handling of missing data on questionnaires

As MMRM is the main analysis, no handling of missing values will be done. Furthermore, at the time of the BDRM no excessive amount of missing data were observed on the questionnaires, so no imputation rules will be put in place.

4.6.2.8. Subgroup analysis

The descriptive analyses of the co-primary endpoints analysis will be repeated by region (North America, Europe) on the FAS.

## 4.7. Pharmacokinetic Evaluations

Blood samples for assay of NT-814 plasma concentrations are collected at each of Weeks 2, 4, 8, and 12. The plasma NT-814 concentrations will be listed by scheduled visit on the Safety Population. Further details of the analysis of PK data will be described in an Exposure-Response Data Analysis Plan, and exposure-response data will be reported in a separate report.

## 4.8. Safety Analyses

Safety analyses will be conducted using the Safety Analysis Set.

4.8.1. Treatment Exposure

Definitions of exposure variables for each drug are provided in Table 3.

**Table 3 Exposure variables definitions**

| **Variables** | **Definitions** |
|---|---|
| Duration of dosing (days) | Last dose of randomised blinded treatment – first dose of randomised blinded treatment +1 |
| Duration of dosing (weeks) | (Last dose of randomised blinded treatment – first dose of randomised blinded treatment +1) / 7 |

| | |
|---|---|
| Study treatment compliance (%) | 100 * (total number of capsules taken) / (total number of capsules planned to be taken)<br><br>Note: number of capsules planned to be received is 4*duration of dosing (in days).<br><br>Total number of capsules taken is calculated using the eCRF pages 'Drug Accountability" and "IMP dispense/administration", based on the fact that each weekly card contains 32 capsules and four weekly cards will be dispensed at each of baseline, week 4 and week 8. Thus, one visit dispensed kit contains 32*4=128 capsules. At subsequent Visits (Weeks 4, 8 and 12) the number of capsules remaining within the 4 weekly cards is recorded and the number of capsules taken = 128 - Number of capsules remaining in the 4 weekly cards.<br><br>In the event that the subject doesn't return some or all of the 4 weekly cards at Week 4, 8 or 12, it will not be possible to determine the number of capsules taken, so protocol medication adherence taken will be set as missing. In the event that more than one kit was dispensed to the subject between 2 visits, in a same way it will not be possible to determine the number of capsules taken, so protocol medication adherence taken will be set as missing. Unscheduled visits will be also included in the calculation if Drug Accountability eCRF Form is completed. |
| Cumulative NT-814 exposure (mg) | It is assumed that on a day the subject takes the 4 planned capsules:<br><br>In Placebo group, Cumulative dose = 0<br><br>In NT-814 40mg group, Cumulative dose = (number of capsules taken / 4) * 40 mg.<br><br>In NT-814 80mg group, Cumulative dose = (number of capsules taken / 4) * 80 mg.<br><br>In NT-814 120mg group, Cumulative dose = (number of capsules taken / 4) * 120 mg.<br><br>In NT-814 160mg group, Cumulative dose = (number of capsules taken /4) * 160 mg. |

The following information will be tabulated overall and by treatment:

- Summary of randomised drug exposure: Duration of dosing will be summarized quantitatively and qualitatively for the category (≤4 weeks, >4-≤8 weeks, >8-≤12 weeks, >12 weeks)
- Summary of cumulative NT-814 exposure (in mg)

- Summary of randomised study treatment compliance: Compliance will be summarized quantitatively and qualitatively for the category (≤ 50%, >50 to ≤ 80%, > 80%).

Listing of IMP dispense/administration and drug accountability data will be provided for the Safety Analysis Set. Placebo run in exposure data will be only provided in a listing for the Screen Analysis Set.

4.8.2. Adverse Events

Adverse events will be coded using MedDRA Version 21.1 and displayed in tables and listings using SOC and PT.

Analyses of adverse events will be performed for those events that are considered treatment emergent (TEAE), where treatment emergent is defined as any adverse event with the onset date is on or after the date and time of first dosing with randomised study treatment. Any adverse event with an onset date earlier than the first dosing with randomised study treatment will be considered as a pre-treatment adverse event. In case there is any missing or incomplete onset date, the adverse event will be classified as treatment-emergent if the partial adverse event onset date/time information does not indicate that the adverse event started prior to the date and time of first dosing with study treatment. No imputation of adverse event dates/times will be performed.

If a subject experiences more than one AE with the same preferred term, that preferred term will be counted only once within summaries. It will be assigned the highest observed severity and the strongest relationship to study treatment among those events for the tables in which those characteristics are summarised.

An overview of adverse events will be presented and will include the number and percentage of patients with at least one:

- Treatment emergent adverse events
- Treatment emergent adverse events related to IMP
- Serious treatment emergent adverse events
- Treatment emergent adverse events leading to treatment discontinuation
- Treatment emergent adverse events leading to study discontinuation
- Treatment emergent adverse events leading to Death
- Severe treatment emergent adverse events

The number and percentage of subjects with the following adverse events will be presented by SOC and PT:

- Treatment emergent adverse event
- Serious treatment emergent adverse event
- Treatment emergent adverse event related to IMP
- Treatment emergent adverse event leading to treatment discontinuation
- Treatment emergent adverse events leading to study discontinuation

In addition, the treatment emergent adverse events will be summarized by SOC, by PT and by severity and by relationship to study treatment. For a patient with more than one occurrence of the same adverse event in a particular SOC/PT, only the adverse event with the most severe intensity and / or most extreme relationship to the study drug will be considered.

By-subject supportive listings will also be provided for the following:

- Pre-treatment AE;
- Treatment emergent adverse events;
- All Serious adverse events;
- Severe treatment emergent adverse events;
- Treatment emergent adverse events leading to treatment discontinuation;
- Treatment emergent adverse events leading to study discontinuation;
- Treatment emergent adverse events leading to death.

4.8.3. Laboratory Data

Clinical laboratory values will be expressed using conventional SI units.

Laboratory parameters (haematology, biochemistry, bone turnover markers and urinalysis) will be summarised over the scheduled visits by treatment group. The actual value and change from baseline will be summarised for the haematology, biochemistry, urinalysis and bone turn-over parameters based on central laboratory measurements.

Urinalysis categorical parameters will be summarised using the number and percentage of patients within each category.

In addition, for haematology and biochemistry parameters, shift tables from baseline to each post-baseline value will be produced using the low/normal/high classification based on laboratory reference ranges. For some parameters with multiple reference ranges depending on menopausal status or menstrual cycle indicating in a comment variable, the one referenced to post-menopausal will be applied.

All laboratory data will be provided in data listings. Laboratory values outside the reference range will be identified in the subject listings.

A subset listing will be presented for all laboratory values with an overall abnormal clinically significant assessment. A listing for the description of abnormal assessments will be provided.

Urine pregnancy test data will be also presented in a listing.

**Table 4 List of Laboratory Parameters**

| Category | Parameters |
|---|---|
| **Haematology** | red blood cell (RBC) count, white blood cell (WBC) count, haematocrit, haemoglobin, MCV, platelet count and WBC differentials |

| | |
|---|---|
| **Biochemistry** | sodium, potassium, glucose, urea (blood urea nitrogen), creatinine, creatine kinase, albumin, calcium, phosphate, bilirubin (total), alkaline phosphatase, aspartate transaminase (AST), alanine transaminase (ALT), gamma glutamyl transferase (GGT), bicarbonate, magnesium, chloride, total protein, haemoglobin A1c (HbA1c) |
| **Urinalysis** | glucose, bilirubin, ketones, specific gravity, blood, pH, protein, urobilinogen, nitrites, leucocyte esterase, sedimentation |
| **Bone Turnover Markers** | bone-specific alkaline phosphatase (BALP) and procollagen type 1 N-terminal pro-peptide (P1NP) |

4.8.4. Vital Signs and Physical Examinations

Vital signs parameters include weight (kg), BMI, heart rate (beat/min, (bpm)), systolic and diastolic blood pressure (mmHg), waist circumference (cm) and temperature (°C).

The following conversions will be applied:

Weight (kg) = 0,45359237 * Weight (lb)

Waist circumference (cm) = 2,54 * Waist circumference (Inches)

Temperature (°C) = (Temperature (°F) – 32) / 1.8)

The actual value and change from Baseline will be summarised descriptively by treatment group and time point for vital signs parameters.

By-subject listings of vital sign measurements will be presented. A listing for the description of abnormal assessments will be provided.

Physical examination includes a review of the following body systems: General appearance, Skin, Head, eyes, ears, nose and throat, Respiratory, Cardiovascular, Abdomen (including liver and kidneys), Musculoskeletal and Neurological. All physical examination findings will be presented in a data listing.

4.8.5. Electrocardiogram

ECG interval parameters include RR interval (msec), PR interval (msec), QT interval (msec), QTc interval (msec) and QTcF interval (msec).

QTcF interval will be derived using: QT interval/(RR interval)^1/3

The actual value and change from Baseline will be summarised descriptively by treatment group and time point for ECG parameters.

The proportion of subjects with absolute QTcF values by category below will be summarized by time point for the following categories: ≤450, >450 to ≤480, >480 to ≤500, >500 msec.

Similarly, the proportion of subjects with an increase from baseline in QTcF values will be summarized by time point for the following categories: ≤0, >0 to ≤30, >30 to ≤60, >60 msec.

ECG Overall interpretation (normal, abnormal clinically and not clinically significant results) will be summarized at baseline and each study visit.

All ECG data for each subject will be provided in data listings. A listing for the description of abnormal assessments will be provided.

### 4.8.6. Electronic Columbia Suicide Severity Rating Scale (eC-SSRS)

The Columbia Suicide Severity Rating Scale, or C-SSRS, is a rating scale created to evaluate suicidality in adults and children over the age of 12. It rates an individual's degree of suicidal ideation on a scale, ranging from "wish to be dead" to "active suicidal ideation with specific plan and intent".

Scoring:

The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.

Category 1 – Wish to be Dead
Category 2 – Non-specific Active Suicidal Thoughts
Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5 – Active Suicidal Ideation with Specific Plan and Intent
Category 6 – Preparatory Acts or Behavior
Category 7 – Aborted Attempt
Category 8 – Interrupted Attempt
Category 9 – Actual Attempt (non-fatal)

Composite endpoints based on the above categories are defined below:

**Suicidal ideation**: A “yes” answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.

**Suicidal behavior**: A “yes” answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-9) on the C-SSRS.

**Suicidal ideation or behavior**: A “yes” answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-9) on the C-SSRS.

**Suicidal Ideation Score**: The maximum suicidal ideation category (1-5 on the CSSRS) present at the assessment. Score will be assigned to 0 if no ideation is present. This score has a range of 0 to 5.

Analysis:

The 9 C-SSRS categories, Suicidal ideation, Suicidal behavior and Suicidal ideation or behavior will be summarised descriptively by treatment group and time point. For each item and time point, the number of patients with a response 'yes' will be presented.

Shift tables from baseline to each post-baseline visit will be presented for the suicidal ideation categories. Counts and percentages will be displayed.

### 4.8.7. Concomitant Medications

Prior medications are those the patient used prior to first day of randomised treatment, so with a stop date and time before randomised study treatment start date and time.

Concomitant medications are those the patient used on or after first day and time of treatment. No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a concomitant medication.

Concomitant medications will be coded using the WHO Drug Dictionary Version Sep2018 B3 Global. The number and percentages of patients with at least one medication will be tabulated by Anatomic Therapeutic Class (ATC level 2) and preferred term. ATC classes will be sorted by descending order of frequency in the total column and the same rule applies for preferred terms within each ATC class. Previous therapies and concomitant therapies will be summarized separately.

The use of prior and concomitant medications will be included in by-subject data listing.

Concomitant non-drug treatments will be coding using MedDRA Version 21.1, and only included in by-subject data listing.

# 5. CHANGES TO PLANNED ANALYSES

As of this date, there have been no changes between the protocol-defined statistical analyses and those presented in this statistical plan.

# 6. REFERENCES

[1] COMPASS 1.1 (E) User Manual (2012), Cytel Inc., Cambridge, MA, USA.

[2] Ivanova, A., Bolognese, J. A., Perevozskaya I. Adaptive dose finding based on t-statistic for dose-response trial. Statistics in Medicine 2008, 27:1581-1592.

# 7. CLINICAL STUDY REPORT APPENDICES

## 7.1. Statistical Tables to be Generated

Sample tables and numbering are provided below.

The ones highlighted in yellow are the ones that will be delivered at each IA.

*Disposition/Demographics/Baseline/Exposure (CSR Table Section 14.1)*

| | |
|---|---|
| Table 14.1.1.1 | Subject Enrolment and Disposition (Screen Analysis Set) |
| Table 14.1.1.2 | Treatment and Study Completion (Randomised Analysis Set) |
| Table 14.1.1.3 | Overview of Analysis Sets (Randomised Analysis Set) |
| Table 14.1.1.4 | Enrolment by Country and by Site (Safety Analysis Set) |
| Table 14.1.1.5 | Relevant protocol Deviations (Randomised Analysis Set) |
| Table 14.1.2 | Demographic and Baseline Characteristics (Safety Analysis Set) |
| Table 14.1.3 | Baseline Menopause Characteristics (Safety Analysis Set) |
| Table 14.1.4 | Medical History (Safety Analysis Set) |
| Table 14.1.5 | Exposure to Randomised Study Treatment (Safety Analysis Set) |

*Efficacy/Pharmacokinetic Results (CSR Table Section 14.2)*

| | |
|---|---|
| Table 14.2.1.1A | Mean Daily Frequency of Moderate and Severe Hot Flushes and Change From Baseline by Week (Full Analysis Set) |
| Table 14.2.1.1B | Mean Daily Frequency of Moderate and Severe Hot Flushes and Change From Baseline by Week (PP Analysis Set) |
| Table 14.2.1.1C | Mean Daily Frequency of Moderate and Severe Hot Flushes and Change From Baseline by Week (Modified Full Analysis Set) |
| Table 14.2.1.2A | Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline - MMRM Analysis (Full Analysis Set) |
| Table 14.2.1.2B | Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline - MMRM Analysis (PP Analysis Set) |

| | |
|---|---|
| Table 14.2.10.2 | Insomnia Severity Index Score Change From Baseline - MMRM Analysis (Full Analysis Set) |
| Table 14.2.11.1 | HFRDIS Scores and Change From Baseline by Visit (Full Analysis Set) |
| Table 14.2.11.2 | HFRDIS Scores Change From Baseline - MMRM Analysis (Full Analysis Set) |
| Table 14.2.12.1 | MenQoL-I Scores and Change From Baseline by Visit (Full Analysis Set) |
| Table 14.2.12.2 | MenQoL-I Scores Change From Baseline - MMRM Analysis (Full Analysis Set) |
| Table 14.2.13.1 | Beck Depression Inventory II Score and Change From Baseline by Visit (Full Analysis Set) |
| Table 14.2.13.2 | Beck Depression Inventory II Score Change From Baseline - MMRM Analysis (Full Analysis Set) |
| Table 14.2.14.1 | Mean Daily Frequency of Moderate and Severe Hot Flushes and Change From Baseline by Week and by Region (Full Analysis Set) |
| Table 14.2.14.2 | Mean Weekly Severity of Moderate and Severe Hot Flushes and Change From Baseline by Week and by Region (Full Analysis Set) |

*Displays of Adverse Events (CSR Section 14.3.1)*

| | |
|---|---|
| Table 14.3.1.1 | Overview of Treatment-Emergent Adverse Events (TEAE) (Safety Analysis Set) |
| Table 14.3.1.2 | Treatment-Emergent Adverse Events (TEAE) by System Organ Class and Preferred Term (Safety Analysis Set) |
| Table 14.3.1.3 | Serious Treatment-Emergent Adverse Events (TEAE) by System Organ Class and Preferred Term (Safety Analysis Set) |
| Table 14.3.1.4 | Drug-Related Treatment-Emergent Adverse Events (TEAE) by System Organ Class and Preferred Term (Safety Analysis Set) |
| Table 14.3.1.5 | Treatment-Emergent Adverse Events (TEAE) Leading to Treatment Discontinuation by System Organ Class and Preferred Term (Safety Analysis Set) |

| | |
|---|---|
| Table 14.3.1.6 | Treatment-Emergent Adverse Events (TEAE) Leading to Study Discontinuation by System Organ Class and Preferred Term (Safety Analysis Set) |
| Table 14.3.1.7 | Treatment-Emergent Adverse Events (TEAE) by System Organ Class, by Preferred Term and by Severity (Safety Analysis Set) |
| Table 14.3.1.8 | Treatment-Emergent Adverse Events (TEAE) by System Organ Class, by Preferred Term and by Relationship to Study Treatment (Safety Analysis Set) |

*Other Safety Information (CSR Section 14.4)*

| | |
|---|---|
| Table 14.4.1.1 | Clinical Laboratories: Hematology Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.1.2 | Clinical Laboratories: Biochemistry Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.1.3 | Clinical Laboratories: Bone Turnover Markers Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.1.4 | Clinical Laboratories: Urinalysis Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.1.5 | Clinical Laboratories: Urinalysis Categorical Actual Values Over Time on Study (Safety Analysis Set) |
| Table 14.4.2.1 | Clinical Laboratories: Shifts from Baseline Versus Each Post-Baseline Visit for Hematology (Safety Analysis Set) |
| Table 14.4.2.2 | Clinical Laboratories: Shifts from Baseline Versus Each Post-Baseline Visit for Biochemistry (Safety Analysis Set) |
| Table 14.4.3 | Vital Signs: Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.4.1 | Electrocardiogram Interval Results: Actual Values and Changes Over Time on Study (Safety Analysis Set) |
| Table 14.4.4.2 | Electrocardiogram Results: QTcF Categorical Actual Values and Changes Over Time on Study (Safety Analysis Set) |

| | |
|---|---|
| Table 14.4.4.3 | Summary of ECG Overall Interpretation by Visit (Safety Analysis Set) |
| Table 14.4.5.1 | Electronic Columbia Suicide Severity Rating Scale Categories Values Over Time on Study (Safety Analysis Set) |
| Table 14.4.5.2 | Electronic Columbia Suicide Severity Rating Scale: Shifts from Baseline Versus Post-Baseline for Suicidal Ideation Scores (Safety Analysis Set) |
| Table 14.4.6 | Prior Medications (Safety Analysis Set) |
| Table 14.4.7 | Concomitant Medications (Safety Analysis Set) |

## 7.2. Data Listings to be Generated

Sample listings and numbering are provided below.

The ones highlighted in yellow are the ones that will be delivered at each IA.

*Study Information (CSR Appendix 16.1)*

| | |
|---|---|
| Listing 16.1.6 | Listing of Patients Receiving Test Drug(s)/Investigational Product(s) from Specific Batches, Where More than One Batch Was Used (Safety analysis Set) |
| Listing 16.1.7 | Randomization Scheme and Codes (Safety Analysis Set)) |

*Discontinued Subjects (CSR Appendix 16.2.1)*

| | |
|---|---|
| Listing 16.2.1.1 | Screen Failure (Screen Analysis Set) |
| Listing 16.2.1.2 | Subject Disposition and Study Termination Information (Randomised Analysis Set) |
| Listing 16.2.1.3 | Study and Visit dates (Screen Analysis Set) |

*Protocol Deviations (CSR Appendix 16.2.2)*

| | |
|---|---|
| Listing 16.2.2.1 | Inclusion/Exclusion Criteria Not Met (Randomised Analysis Set) |
| Listing 16.2.2.2 | Protocol Deviation (Safety Analysis Set) |

*Subjects Excluded from Efficacy Analysis (CSR Appendix 16.2.3)*

| | |
|---|---|
| Listing 16.2.3.1 | Subjects Excluded from Efficacy Analysis (Safety Analysis Set) |

*Demographics Data (CSR Appendix 16.2.4)*

| | |
|---|---|
| Listing 16.2.4.1 | Demographic and Baseline Information (Safety Analysis Set) |
| Listing 16.2.4.2 | Baseline Menopause Characteristics (Safety Analysis Set) |
| Listing 16.2.4.3 | Medical History (Safety Analysis Set) |

*Compliance and/or Drug Concentration Data (CSR Appendix 16.2.5)*

| | |
|---|---|
| Listing 16.2.5.1 | Dosing Information for Placebo Run-in Period (Screen Analysis Set) |
| Listing 16.2.5.2 | Dosing Information for Randomised Study Treatment (Safety Analysis Set) |
| Listing 16.2.5.3 | Exposure to Randomised Study Treatment (Safety Analysis Set) |
| Listing 16.2.5.4 | Orvepitant Concentration Data (Safety Analysis Set) |

*Individual Efficacy Response Data (CSR Appendix 16.2.6)*

| | |
|---|---|
| Listing 16.2.6.1.1 | Hot Flushes eDiary Data (Safety Analysis Set) |
| Listing 16.2.6.1.2 | Hot Flushes Endpoints (Safety Analysis Set) |
| Listing 16.2.6.2.1a | Pittsburgh Sleep Quality Index Individual Items - Questionnaire Key |
| Listing 16.2.6.2.1b | Pittsburgh Sleep Quality Index Individual Items - Questionnaire Responses (Safety Analysis Set) |
| Listing 16.2.6.2.2 | Pittsburgh Sleep Quality Index Scores (Safety Analysis Set) |
| Listing 16.2.6.3.1a | Insomnia Severity Index Individual Items - Questionnaire Key |
| Listing 16.2.6.3.1b | Insomnia Severity Index Individual Items - Questionnaire Responses (Safety Analysis Set) |
| Listing 16.2.6.4.1a | HFRDIS Individual Items - Questionnaire Key (Safety Analysis Set) |
| Listing 16.2.6.4.1b | HFRDIS Individual Items and Scores - Questionnaire Responses (Safety Analysis Set) |
| Listing 16.2.6.5.1a | MenQoL-I Individual Items - Questionnaire Key |

Listing 16.2.6.5.1b MenQoL-I Individual Items - Questionnaire Responses (Safety Analysis Set)

Listing 16.2.6.5.2 MenQoL-I Scores (Safety Analysis Set)

Listing 16.2.6.6.1a Beck Depression Inventory II Individual Items - Questionnaire Key

Listing 16.2.6.6.1b Beck Depression Inventory II Individual Items - Questionnaire Responses (Safety Analysis Set)

Listing 16.2.6.6.2 Beck Depression Inventory II Score (Safety Analysis Set)

*Adverse Event Listings (each subject) (CSR Appendix 16.2.7)*

Listing 16.2.7.1 All Treatment-Emergent Adverse Events (Safety Analysis Set)

Listing 16.2.7.2 All Serious Adverse Events (Safety Analysis Set)

Listing 16.2.7.3 All Severe Treatment-Emergent Adverse Events (Safety Analysis Set)

Listing 16.2.7.4 All Treatment-Emergent Adverse Events Leading to Treatment Discontinuation (Safety Analysis Set)

Listing 16.2.7.5 All Treatment-Emergent Adverse Events Leading to Study Discontinuation (Safety Analysis Set)

Listing 16.2.7.6 All Treatment-Emergent Adverse Events Leading to Death (Safety Analysis Set)

Listing 16.2.7.7 Pre-Treatment Adverse Events (Safety Analysis Set)

*Listing of Individual Laboratory Measurements by Subject (CSR Appendix 16.2.8)*

Listing 16.2.8.1 Laboratory Results: Hematology (Safety Analysis Set)

Listing 16.2.8.2 Laboratory Results: Chemistry (Safety Analysis Set)

Listing 16.2.8.3 Laboratory Results: Urinalysis (Safety Analysis Set)

Listing 16.2.8.4 Laboratory Results: Bone Turnover Markers (Safety Analysis Set)

Listing 16.2.8.5 Overall Abnormal Clinically Significant Assessment for Laboratory Parameters (Safety Analysis Set)

| | |
|---|---|
| Listing 16.2.8.6 | Description of Abnormal Assessments for Laboratory Parameters (Safety Analysis Set) |
| Listing 16.2.8.7 | Urine Pregnancy Test (Safety Analysis Set) |

*Other Safety Data (CSR Appendix 16.2.9)*

| | |
|---|---|
| Listing 16.2.9.1.1 | Vital Signs (Safety Analysis Set) |
| Listing 16.2.9.1.2 | Description of Abnormal Assessments for Vital Signs (Safety Analysis Set) |
| Listing 16.2.9.2 | Physical Examination (Safety Analysis Set) |
| Listing 16.2.9.3.1 | Electrocardiograms (Safety Analysis Set) |
| Listing 16.2.9.3.2 | Description of Abnormal Assessments for Electrocardiograms (Safety Analysis Set) |
| Listing 16.2.9.4 | C-SSRS (Safety Analysis Set) |
| Listing 16.2.9.5 | Prior Medications (Safety Analysis Set) |
| Listing 16.2.9.6 | Concomitant Medications (Safety Analysis Set) |
| Listing 16.2.9.7 | Concomitant Non-Drug Treatments (Safety Analysis Set) |

## 7.3. Data Figures to be Generated

Sample figures and numbering are provided below.

*Efficacy/Pharmacokinetic Results (CSR Table Section 14.2)*

| | |
|---|---|
| Figure 14.2.1.1A | Change From Baseline Over Time in Mean Daily Frequency of Moderate and Severe Hot Flushes by Treatment Group (Full Analysis Set) |
| Figure 14.2.1.1B | Change From Baseline Over Time in Mean Daily Frequency of Moderate and Severe Hot Flushes by Treatment Group (PP Analysis Set) |
| Figure 14.2.1.2A | Change From Baseline in Mean Daily Frequency of Moderate and Severe Hot Flushes at Week 4 and 12 by Treatment Group (Full Analysis Set) |