CLINICAL STUDY PROTOCOL

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

The "SWITCH-1" Study

IND No:	129492
EudraCT No:	2018-002763-26
ClinicalTrials.gov No:	NCT03596762
Protocol No.:	814-PM-02
Version No:	Final 2.0
Date of Protocol:	07 Feb 2019

Sponsor: NeRRe Therapeutics Ltd,
SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage,
SG1 2FX
United Kingdom

SAE Reporting: Emas Pharma Pharmacovigilance
Tel.: +44 1462 424416
Fax: +44 1462 600456
Email: Drug.Safety@bionical-emas.com

This study will be conducted in compliance with Good Clinical Practice (GCP), the General Principles of the Declaration of Helsinki (with amendments), and in accordance with local legal and regulatory requirements.

1. PROTOCOL SYNOPSIS

PROTOCOL 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**
PROTOCOL NUMBER	814-PM-02
CHIEF INVESTIGATOR	Dr James A. Simon, MD, CCD, NCMP, IF, FACOG Clinical Professor, George Washington University IntimMedicine Specialists 1850 M Street, NW #450 Washington, DC 20036, United States
SPONSOR	NeRRe Therapeutics Ltd
INVESTIGATIONAL MEDICINAL PRODUCT	NT-814
PHASE OF DEVELOPMENT	2b
INDICATION	Post-menopausal vasomotor symptoms
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 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. Subjects who provide informed consent will be screened for participation in the study. After giving informed consent, but before Screening Visit 2 (i.e. before starting the placebo run-in period), subjects will be withdrawn from prohibited concomitant medications. At the first visit (Screening Visit 1), they will be provided with a paper diary, which they will be asked to complete once daily (prior to retiring to bed) for 7 days, recording their recall of the total number of hot flushes (also known as hot flashes) over the previous 24 hours, and the severity of each of those hot flushes. At the second visit (Screening Visit 2), the investigator or designee will review the paper diary to determine if the subject can continue in the study. To continue, subjects must have completed the paper diary for at least 6 days and recorded an average of at least 8 moderate or severe hot flushes per day over the last 5 days that the paper diary was completed. Subjects who continue to satisfy all the initial selection criteria at Screening Visit 2 will commence placebo treatment on a single-blind (subject only) basis and be provided with an electronic diary (eDiary) which they will be asked to complete twice a day for 14 ($\pm$2) days during the remaining screening/baseline period. The eDiary will record the subject's recall of the number of hot flushes during the day (recorded just prior to retiring to bed) and overnight (recorded on wakening). The severity of each hot flush will also be recorded. To proceed to randomisation, the subject will be required to have completed the eDiary for at least 9 days and to have recorded an average of at least 7 moderate or severe hot flushes per day over the last 7 days that the eDiary was completed. At the baseline visit (Day 1), subjects who comply with all selection criteria will be enrolled into the 12-week double-blind portion of the study and be randomised to one of the NT-814 doses or placebo. Subjects who do not meet the eDiary criteria will not continue into the study (they will be screen failures).

	Randomised subjects will undergo further clinic visits at Week 2, Week 4, Week 8, Week 12 (end of treatment visit) and Week 16 (end of study) with regular assessments of compliance with study medication, compliance with the eDiary completion as well as assessments of concomitant medications, mental well-being, quality of life (QoL), sleep, and safety. A treatment kit will be dispensed at Screening Visit 2 (for the placebo run-in), the Baseline Visit, Week 4 and Week 8. A final follow-up safety assessment will be performed at Week 16, 4-weeks after stopping study medication. Blood samples will be collected for pharmacokinetic (PK) analysis for all subjects at Weeks 2, 4, 8, and 12.
STUDY OBJECTIVES	
Primary	• To evaluate the efficacy of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, in reducing the frequency and severity of hot flushes. • To assess the safety and tolerability of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, in subjects with post-menopausal symptoms.
Secondary	• To evaluate the effect of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, on mental well-being, quality of life and measures of sleep in subjects with post-menopausal symptoms. • To evaluate the dose-response relationship of 40 mg, 80 mg, 120 mg and 160 mg NT-814 in reducing hot flush frequency and severity.
Pharmacokinetics	• To evaluate the exposure-response relationship of NT-814 using population pharmacokinetics (PK) with sparse sampling.
Pharmacodynamics	• Not applicable
EFFICACY ASSESSMENTS AND ENDPOINTS	
EFFICACY ASSESSMENTS	EFFICACY ENDPOINTS
• Number of hot flushes recorded in the eDiary	**Co-Primary Efficacy Endpoints:** • Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Week 4

• Severity of each hot flush recorded on a 3-point scale (mild, moderate or severe)	• Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Week 12 • Mean change from baseline in severity of moderate and severe hot flushes from baseline to Week 4 • Mean change from baseline in severity of moderate and severe hot flushes from baseline to Week 12 **Secondary Efficacy Endpoints:** • Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16 • Mean change from baseline in severity of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16 • Mean change from baseline in frequency of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16 • Mean change from baseline in severity of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16 • Mean change from baseline in the Hot Flush Score (frequency x severity) at Weeks 1, 2, 4, 8, 12 and 16 • Responder analyses (proportion of subjects with $\geq$50% and $\geq$80% reductions from baseline in hot flush frequency at Week 12)
• Number of night time awakenings in the diary either due to a hot flush or unrelated to flushing	• Mean change from baseline in number of night time awakenings secondary to hot flush at Weeks 1, 2, 4, 8, 12 and 16 • Mean change from baseline in number of all night time awakenings at Weeks 1, 2, 4, 8, 12 and 16
• Pittsburgh Sleep Quality Index, completed at Baseline and Weeks 4, 8, 12 and 16	• Change from baseline in the Global and individual domain scores at Weeks 4, 8, 12 and 16
• Insomnia Severity Index completed at Baseline and Weeks	• Change from baseline in the Insomnia Severity Index score at Weeks 4, 8, 12 and 16

4, 8, 12 and 16	
• Hot Flush Related Daily Interference Scale (HFRDIS) completed at all visits except Screening	• Change from baseline in the HFRDIS scores at Weeks 2, 4, 8, 12 and 16
• Menopause-Specific Quality of Life Intervention Version (MenQoL-I) completed at Baseline and Weeks 4, 8, 12 and 16	• Change from baseline in the MenQoL-I scores at Weeks 4, 8, 12 and 16
• Beck Depression Inventory (Version II) completed at all visits except Screening	• Change from baseline in the Beck Depression Inventory II scores at Weeks 2, 4, 8, 12 and 16
SAFETY ASSESSMENTS	SAFETY ENDPOINTS
• Adverse events recorded throughout the study (from Screening Visit 2 to Week 16)	• Number of treatment emergent adverse events • Number of treatment emergent serious adverse events • Number of treatment emergent adverse events resulting in treatment discontinuation • Severity of treatment emergent adverse events • Number of treatment emergent related adverse events
• Physical Examination recorded at Screening Visit 2, Baseline and Week 16	• Treatment emergent changes in physical examination will be recorded as adverse events
• Weight recorded at each visit (except Screening Visit 1) and height at Screening Visit 2 • Waist circumference recorded at each visit (except Screening visits 1 and 2)	• Change from baseline in weight at Weeks 2, 4, 8, 12 and 16 • Change from baseline in body mass index at Weeks 2, 4, 8, 12 and 16 • Change from baseline in waist circumference at Weeks 2, 4, 8, 12 and 16
• Electronic Columbia Suicide Severity Rating Scale (eC-SSRS) recorded at	• Change from baseline in eC-SSRS at Weeks 4, 12 and 16

Baseline and Weeks 4, 12 and 16	
• Vital Signs (pulse rate, systolic and diastolic blood pressure, temperature) recorded at each visit (except Screening Visit 1)	• Change from baseline at Weeks 2, 4, 8, 12 and 16 in each vital sign: - Systolic blood pressure - Diastolic blood pressure - Pulse rate - Temperature
• Haematology, clinical chemistry (including HbA1c) and urinalysis parameters, collected at each visit (except Screening Visit 1 and also Week 8 for haematology and urinalysis)	• Change from baseline at Weeks 2, 4, 12 and 16 in each haematology and urinalysis parameter • Change from baseline at Weeks 2, 4, 8, 12 and 16 for each clinical chemistry parameter
• 12-lead ECG, recorded at each visit (except Screening Visit 1)	• Proportion of subjects with clinically significant abnormal ECG findings at each visit • Proportion of subjects with non-significant abnormal findings at each visit • Change from baseline at Weeks 2, 4, 8, 12 and 16 in ECG intervals: - PR - QT, QTc and QTcF - RR • Proportion of subjects with maximum absolute QTcF values by category at each visit - $\leq$450, >450 to $\leq$480, >480 to $\leq$500, >500 msec • Proportion of subjects with maximum change from baseline in QTcF values by category at Weeks 2, 4, 8, 12 and 16 - $\leq$0, >0 to $\leq$30, >30 to $\leq$60, >60 msec
• Plasma bone turnover markers measured at Baseline and Weeks 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 pro-peptide (P1NP)

	Baseline and week 12 and 16 visits should be conducted at similar times of day to reduce diurnal variation in these markers.
PHARMACOKINETIC ASSESSMENTS	PHARMACOKINETIC ENDPOINTS
• Single samples for NT-814 plasma concentration determination at Weeks 2, 4, 8 and 12	• Exposure-response modelling will be undertaken on a number of efficacy and safety endpoints on an exploratory basis.
ELIGIBILITY CRITERIA	
INCLUSION CRITERIA	1. Females aged 40 to 65 years, inclusive, at Screening Visit 1 2. Able to understand and comply with the requirements of the study and give informed consent 3. Postmenopausal, defined as: (i) at least 12 months of spontaneous amenorrhea, or (ii) at least 6 months of spontaneous amenorrhea with serum FSH levels > 40 mIU/ml and a serum oestradiol concentration of < 30 pg/mL, or (iii) at least 6 weeks postsurgical bilateral oophorectomy with or without hysterectomy[1],[2] 4. Body mass index between 18 and 38 kg/m^2, inclusive, at Screening Visit 2 5. Negative urinary pregnancy test at Screening Visit 2 6. In good general health, in the opinion of the investigator, based on the medical history, physical examination, 12-lead ECG, vital signs and clinical laboratory tests assessed at Screening Visit 2 7. Subject has completed the paper diary for at least 6 days between Screening Visits 1 and 2 and has recorded an average of at least 8 moderate or severe hot flushes per day (including night-time) over the last 5 days that the paper diary was completed (assessed at Screening Visit 2) 8. Subject has completed the eDiary for at least 9 days between Screening Visit 2 and Day 1 and has recorded an average of at least 7 moderate or severe hot flushes per day (including night-time) over the last 7 days that the eDiary was completed (assessed at the Baseline Visit)

[1] Durations are relative to screening visit 1

[2] Subjects who do not clearly fall into 1 of the 3 defined categories, but who are hormonally post-menopausal (serum FSH levels > 40 mIU/ml and a serum oestradiol concentration of < 30 pg/mL) and of an appropriate age, may also be considered for inclusion.

EXCLUSION CRITERIA	1. Have used or unwilling to wash-out use of any of the following hormonal therapies for the periods stated prior to Screening Visit 2: - ≥1 week for vaginal hormonal products (rings, creams, gels and including DHEA or analogues thereof) - ≥4 weeks for transdermal oestrogen alone or oestrogen/progestin products - ≥8 weeks for oral oestrogen (including selective oestrogen receptor modulators) and/or progestin therapy - ≥8 weeks for intrauterine progestin therapy - ≥3 months for progestin implants and oestrogen alone injectable drug therapy - ≥6 months for oestrogen pellet therapy or progestin injectable drug therapy 2. The use of non-hormonal prescription (eg paroxetine, other anti-depressants, alpha agonists [eg clonidine], methyldopa, gabapentin, pregabalin, medicinal cannabis) or over the counter/herbal treatments for treatment of menopausal symptoms is not allowed throughout the study. Subjects must have discontinued these drugs at least 28 days prior to Screening Visit 2. Subjects may, however, be permitted to continue to use these drugs if the dose has been stable for at least 4 weeks and they have been prescribed solely for the management of another disorder (e.g. neuropathic pain, depression). 3. Inability to comply with the use of prohibited medications as described below (See Appendix A for more details): i. Use of digoxin is not allowed from Screening Visit 2 until 1 week after the last dose of IMP ii. Use of known CYP3A4 substrates with a narrow therapeutic range is not allowed from Screening Visit 2 until 1 week after the last dose of IMP iii. Use of strong or moderate inhibitors of CYP3A4 is not allowed from Screening Visit 2 until 1 week after the last dose of IMP iv. Use of moderate or strong inducers of CYP3A4 is not allowed from Screening Visit 2 until Week 12 v. Use of known P-glycoprotein inhibitors is not allowed from Screening Visit 2 until 1

	week after the last dose of IMP 4. Any prior or ongoing history of clinically relevant drug or alcohol abuse within 12 months of Screening Visit 1 5. Any clinically significant prior or ongoing history of arrhythmias, either determined through clinical history or on ECG evaluation. 6. Any clinically significant abnormal laboratory test result(s), measured at Screening Visit 2. 7. Any active ongoing condition that could have caused difficulty in interpreting vasomotor symptoms such as: infection that could have caused pyrexia, pheochromocytoma, hyperthyroidism, carcinoid syndrome, alcohol abuse. 8. Current history or previous (within the past 5 years) history of any malignancy (except basal and squamous cell skin tumours). 9. Uncontrolled hypertension[3] 10. A history of hyperthyroidism or hypothyroidism. Treated hypothyroidism with normal thyroid function test results at Screening Visit 2 and a stable (for ≥3 months before Screening Visit 2) dose of replacement therapy is acceptable. 11. Known hypersensitivity to NT-814 or any of the excipients in the formulation. 12. Concurrent (or within the 2 months prior to Screening Visit 1) participation in a clinical study with an investigational medicinal product. 13. Concurrent (or within the 1 month prior to Screening Visit 1) participation in an interventional clinical study. 14. Previous participation in a clinical study with NT-814. 15. Dependent on the investigator, the contract research organisation(s) or Sponsor for education or employment. 16. Any unexplained post-menopausal bleeding. 17. Abnormal findings on mammogram or subject has not undergone a mammogram within the guidelines recommended by applicable national authorities (eg United Kingdom National Health Service, United

[3] As a guide, a blood pressure of ≥140 mmHg (systolic) or ≥90 mmHg (diastolic) will usually require further evaluation. This may include a repeat after a period of rest/relaxation.

	States Preventative Services Taskforce, Canadian Task Force on Preventative Healthcare)[4]
NT-814 FORMULATION/DOSE	NT-814 40 mg soft gel capsules Dose IMP Supply 40 mg 1 x NT-814 capsule, 3 x placebo capsules 80 mg 2 x NT-814 capsules, 2 x placebo capsules 120 mg 3 x NT-814 capsules, 1 x placebo capsule 160 mg 4 x NT-814 capsules
REFERENCE FORMULATION/DOSE	Matching placebo soft gel capsules 4 capsules
ROUTE OF ADMINISTRATION	Oral
DURATION/FREQUENCY OF TREATMENT	Once daily dosing in the evening for 12 weeks. All subjects will also receive placebo capsules for the last 2 weeks of the screening/baseline period.
PLANNED SAMPLE SIZE	For the efficacy objectives of the study, 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 hot flushes 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. The actual sample size will be determined by the adaptive dose-finding design algorithm outcomes. In addition, the initial actual sample size is increased by 10% to allow for subjects who withdraw prematurely and so the initial sample size will be 165 subjects. The sample size allows for 95% power since each of the four co-primary endpoints are required to demonstrate statistical significance to yield a firm conclusion of beneficial treatment effect on hot flushes. Conservatively, assuming independence among the four endpoints, this would yield approximately 81% power overall for achieving statistical significance for each of

[4] Subjects who have not had a mammogram within the guidelines may have one performed as part of the screening process

	the four endpoints. Thus, overall power for the trial is estimated to be at least 81%; the actual power depends on the unknown magnitudes of correlations among the co-primary endpoints.
STATISTICAL METHODS	The Safety Analysis Set consists of 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. The Full Analysis Set (FAS) consists of all randomised subjects who received at least one dose of double-blind study drug, irrespective of treatment received, and have a 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. The Per Protocol analysis set consists of all subjects in the FAS excluding those identified as having relevant protocol deviations. The Exposure-Response Set consists of all subjects who received at least one dose of double-blind study drug and for whom the PK data are considered sufficient. All statistical hypothesis tests and confidence intervals will be two sided, using a type I error rate of 0.05. No adjustment for multiple comparisons will be used for this Phase 2 study. The co-primary endpoints are the change from baseline in the mean daily frequencies of moderate and severe hot flushes in the 7 days before the Week 4 and Week 12 visits, and the mean severity of moderate and severe hot flushes in the 7 days before the Week 4 and Week 12 visits. The baseline assessment will be the last 7 days of the baseline eDiary completion period. Absolute and changes from baseline in the mean daily frequencies and average hot flush severity will be summarised by treatment group. The change from baseline endpoint will be analysed by mixed model repeated measures. Pairwise statistical comparisons are planned for each NT-814 dose group (40 mg, 80 mg, 120 mg and 160 mg) versus placebo, presenting the adjusted mean treatment difference with its corresponding 95% confidence interval. Secondary efficacy endpoints will be analysed in a similar fashion, except that the proportion of responders, defined as a reduction of >=50% points and >=80% points on the weekly average frequency of hot flushes will be analysed by logistic regression.

	The safety analysis set will be used for all presentations of safety endpoints. No statistical testing will be used to compare treatment groups for different safety endpoints. Safety data will be summarised descriptively for each treatment group. All adverse events will be coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA). Treatment-emergent adverse events (TEAE) will be presented within summary presentations, by MedDRA system organ class (SOC), preferred term and treatment group. Blood samples for assay of NT-814 plasma concentrations are collected periodically. The plasma NT-814 concentrations will be listed by visit. 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.

2. TABLE OF CONTENTS

3. LIST OF ABBREVIATIONS

AE	Adverse Event
ADR	Adverse Drug Reaction
ALT	Alanine amino transferase / alanine transaminase
ANCOVA	Analysis of covariance
AST	Aspartate amino transferase / aspartate transaminase
BALP	Bone-specific alkaline phosphatase
BDI-II	Beck Depression Inventory – Version II
BDRM	Blinded Data Review Meeting
BP	Blood Pressure
CHMP	Committee for Medicinal Products for Human Use
eCRF	Electronic Case Report Form
C_{max}	Maximum plasma concentration
CSR	Clinical Study Report
eC-SSRS	Electronic Columbia Suicide Severity Rating Scale
CRO	Contract Research Organization
CSR	Clinical Study Report
CV%	Coefficients of variation
CYP3A4	Cytochrome P450 3A4
DHEA	Dehydroepiandrosterone
DR	Dose Response
DRC	Data Review Committee
ECG	Electrocardiogram
eDiary	Electronic Diary
ER	Exposure-Response
FAS	Full Analysis Set
FDA	Food and Drug Administration
FSH	Follicle Stimulating Hormone
GCP	Good Clinical Practice
GGT	Gamma Glutamyl Transferase
GnRH	Gonadotropin Releasing Hormone
H0	Null Hypothesis
H1	Alternative Hypothesis
HbA1c	Haemoglobin A1c
HFRDIS	Hot Flash Related Daily Interference Scale
hERG	Human ether-a-go-go
HF	Hot Flush / Hot Flash
HRT	Hormone Replacement Therapy
IA	Interim Analysis
IB	Investigators Brochure
ICF	Informed Consent Form
ICH	International Conference for Harmonization
IEC	Independent Ethics Committee
IMP	Investigational Medicinal Product
IRB	Institutional Review Board
IM	Intramuscular
ISI	Insomnia Severity Index
IVR	Interactive Voice Response

IWRS	Interactive Web Response Services
KaNDy	KaNDy Therapeutics Ltd
LH	Luteinising hormone
MCV	Mean Corpuscular Volume
MeDRA	Medical Dictionary for Regulatory Activities
MenQoL-I	Menopause-specific Quality-of-Life Questionnaire Intervention Version
NeRRe	NeRRe Therapeutics Ltd
NK	Neurokinin
NTA	Night-time awakening
PET	Positron Emission Tomography
P1NP	Procollagen type 1 N-terminal pro-peptide
PIS	Participant Information Sheet
PK	Pharmacokinetic
PM	Post-menopausal
Pop PK	Population PK
PP	Per-Protocol
PSQI	Pittsburgh Sleep Quality Index
QD	One a day
QoL	Quality of Life
RBA	Relative bioavailability
RBC	Red Blood Cell
RO	Receptor occupancy
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SOC	System Organ Class
SOP	Standard Operating Procedure
SP	Substance P
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TMF	Trial Master File
VMS	Vasomotor Symptoms
WBC	White Blood Cell
WHO	World Health Organization

4. INVESTIGATORS AND ADMINISTRATIVE STRUCTURE

Chief Investigator	Dr James A. Simon, MD, CCD, NCMP, IF, FACOG Clinical Professor, George Washington University IntimMedicine Specialists 1850 M Street, NW #450 Washington, DC 20036, United States Email: jsimon@intimmedicine.com
Medical Expert (Sponsor)	Stephen Pawsey MBBS FRCA FFPM NeRRe Therapeutics Ltd SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX, UK Phone: +44 (0) 7827 460726 Email: steve.pawsey@nerretherapeutics.com
Contract Research Organization / Monitors	Pharm Olam 1st Floor, One Station Square, Bracknell, Berkshire RG12 1QB UNITED KINGDOM T: +44 (0) 1344 891121 F: +44 (0) 1344 890335
Statistical Contract Research Organization	Cytel Inc., 675 Massachusetts Avenue Cambridge Massachusetts 02139 United States
Pharmacokinetic Laboratory	Aptuit Center for Drug Discovery and Development Via A. Fleming, 4 37135 Verona, Italy
Pharmacovigilance Provider	Emas Pharma Rukhnoor Mahmood, 63-56 Knowl Piece, Wilbury Way, Hitchin, SG4 0TY, United Kingdom Tel: +44 1462 426488 Fax: +44 1462 424426 Email: rukhnoor.mahmood@bionical-emas.com

5. BACKGROUND INFORMATION

5.1 NT-814

Prior to its acquisition, initially by NeRRe Therapeutics Ltd (NeRRe), and then subsequently by KaNDy Therapeutics Ltd (KaNDy), NT-814's laboratory code name was GSK1144814. The GSK laboratory code may be used in this and other trial related document when referring to the studies conducted by GSK, but NT-814 is used for all development activities conducted by NeRRe as Sponsor. NT-814 is being developed in a number of indications in women's health, including the treatment of post-menopausal symptoms.

KaNDy has delegated all regulatory and quality Sponsor responsibilities to NeRRe.

5.1.1 Pharmacology

In vitro binding and functional studies have demonstrated that NT-814 is a potent and selective antagonist of both human neurokinin-1 (NK_1) and NK_3 tachykinin receptors. NT-814 was active in pharmacodynamic models for the NK_1 receptor (the gerbil foot tapping test) and NK_3 receptor (wet dog shaking behaviour in guinea pigs); both in a dose dependent manner. NT-814 also shows non-sedating anxiolytic-like properties in the human threat test in the marmoset.

The brain penetrant properties of NT-814 were confirmed in the gerbil and guinea pig in which receptor occupancy (RO) in the brain was demonstrated at both the NK_1 and NK_3 receptors in a dose- and concentration-dependent manner; the RO at both receptors was sustained for extended periods post-dose. In a positron emission tomography (PET) study in baboons, NT-814 achieved high NK_1 RO at modest plasma concentrations. These studies demonstrated that NT-814 is an orally active antagonist of brain NK_1 and NK_3 receptors *in vivo*.

NT-814 was shown to be >100-fold selective for the human NK_3 receptor and >300-fold for the human NK_1 receptor, in comparison to 53 other non-tachykinin receptors, ion channels, enzymes and transporters.

There were no clinically relevant findings in respiratory or neurobehavioral safety pharmacology studies at doses up to 100 mg/kg or in a cardiovascular study in cynomolgus monkeys at doses up to 60 mg/kg. NT-814 did not inhibit the human ether-a-go-go (hERG) tail current.

5.1.2 Pre-clinical Safety

A full package of pre-clinical safety studies appropriate to the stage of development has been completed with NT-814. This includes single dose studies in marmoset and monkey (up to 1000 mg/kg) and repeat dose studies of up to 13-weeks in rat (up to 100 mg/kg/day) and cynomolgus monkey (up to 40 mg/kg/day). The safety profile was acceptable and supported clinical evaluation with appropriate monitoring.

Genotoxicity and reproductive toxicity studies have also been undertaken. NT-814 was not genotoxic and there was no evidence of embryofetal toxicity in the rat at doses up to

100 mg/kg/day or in the rabbit at doses up to 140 mg/kg/day. A signal with respect to potential phototoxicity warrants monitoring of subjects for skin reactions to sunlight.

5.1.3 Clinical Experience

To date, five clinical studies in healthy volunteers have been completed with single total daily doses up to 250 mg and repeat doses up to 200 mg total daily dose for 28 days. These studies comprise a single dose escalation pharmacokinetic (PK) and safety study incorporating an assessment of NK_1 RO using PET imaging (Study MNK111321), a repeat-dose PK and safety study with a drug interaction assessment and NK_1 RO evaluation (MNK111587), a single-dose study assessing the PK and effect of food on a tablet formulation (MNK112891), a single dose study assessing whether the psychomotor and cognitive effects of alcohol are exacerbated by NT-814 (MNK113476) and a relative bioavailability and food effect study (814-1-01).

Thirty-seven female and 89 male healthy volunteers received NT-814 in these studies. NT-814 was found to have a good pharmacokinetic profile after both single and repeat doses. The PK was characterized by rapid absorption, modest accumulation ratios and approximately linear increases in exposure with increasing dose. The terminal elimination half-life was estimated to be in the range ~15-20 hours and steady state was achieved within 7 days. NT-814 was well tolerated throughout the dose range. The only potential safety signal identified was a possible effect on cardiac rhythm, although review of the individual arrhythmias reported suggested that they were likely to be coincidental. Non-serious AEs considered possibly related to treatment with NT-814 are headache, diarrhoea and somnolence.

Seventy-six post-menopausal female subjects experiencing troublesome hot flushes participated in a 14 day repeat dose study which assessed the safety, efficacy and PK of NT-814 in this indication (Study RELENT-1). Eighteen subjects received placebo and 56 received NT-814 at doses ranging from 50 mg to 300 mg once daily. All doses were well tolerated, and no safety concerns were identified. Extensive Holter monitoring did not identify any treatment effect on cardiac rhythm. There was clear evidence of a beneficial effect on post-menopausal symptoms, with clinically relevant and statistically significant improvements on both daytime hot flush frequency and severity and night-time awakening at the higher doses (150 and 300 mg) compared to placebo.

The PK profile in post-menopausal females was very similar to that observed in younger male volunteers. However, the hard gel capsule formulation used in the study resulted a high degree of within and between subject variability in exposure. The variability was generally greater at the higher doses, with the coefficients of variation (CV%) approaching 150% for some parameters. In addition to the high variability, the mean exposures observed at doses up to 150 mg were not as great as those expected based on previous studies. Consequently, a new formulation was developed in which NT-814 was dissolved in a liquid lipid vehicle and encapsulated in a soft gel capsule. The kinetics of this soft gel formulation were evaluated in a relative bioavailability study and it was found to have acceptable variability (CV% <45%) and good exposures. Administration of the soft-gel capsule with food was found in this study to reduce the maximum

plasma concentration (C_{max}) of NT-814 quite significantly (by around 70%) but with only a modest reduction (~9%) on overall exposure. This blunting of the C_{max} dose not result in lower concentrations at later time points and the overall exposure is anticipated to be sufficient for efficacy in both the fed and fasted states.

This soft gel formulation is the current development form of NT-814 and will be used in this study.

Further information on the pre-clinical and clinical studies undertaken with NT-814 can be found in the current version of the Investigators Brochure (IB).

5.2 Hot flushes in post-menopausal women and rationale for this study

Over the coming decades, it is estimated that more than 1 billion women worldwide will be older than 50 years. Up to 75% of these women will experience adverse symptoms related to menopausal transition, particularly hot flushes (also called flashes) vasomotor symptoms, sleep disturbances and adverse mood effects that have such a negative impact upon patient daily activities and quality of life[1,2,3]. Hot flushes are transient but recurrent episodes of flushing, perspiration and intense heat sensation that are one of the most common, bothersome and distressing symptoms felt by women during the transition to menopause (peri-menopause) and post-menopause periods[4]. They are the leading cause for seeking medical attention during this particular phase of a woman's life, particularly amongst those experiencing the severest symptoms[5,6]. The hot flushes may also be associated with difficulty in sleeping[7, 8] worsen depressive symptoms and signal the onset or relapse of a major depressive episode[9,10]. When hot flushes occur at night they are known as night sweats and can cause insomnia and fatigue[11,12]. Hot flushes can vary in frequency, with some women experiencing episodes many times a day, and can last up to 10 years after the last menses[13,14]. Current treatment options are limited and include hormone replacement therapy (HRT) which is effective in many women but associated with an increased risk of hormone-dependent cancers, thrombotic risk and cardiovascular adverse effects[15]. HRT can also take several months to improve hot flush symptoms, especially when used at lower doses, and has limited effects on sleep disturbance[16,17]. These and other limitations preclude its use in many women. A low dose of the anti-depressant paroxetine is licensed for use in some countries but has limited efficacy and a range of safety and tolerability concerns that prevent its widespread usage especially in the breast cancer survivor population.

Emerging data show that hot flushes may be treated by targeting the neuroendocrine factors that trigger the symptoms[18]. The menopause is characterized by decreased oestradiol levels due to ovarian failure, and a resultant increase in gonadotropin releasing hormone (GnRH) secretion from the hypothalamus leading to high gonadotropin (luteinising hormone [LH] and follicle stimulating hormone) concentrations. A link between initiation of the LH pulse and hot flushes has been reported[19]. However, a number of studies involving oestrogen administration and withdrawal in patients who do not have elevated LH have shown they still experience hot flushes[20,21,22], as do patients with Kallman's syndrome who lack hypothalamic GnRH neurons, and in whom oestrogen withdrawal will trigger hot flushes[23]. These findings show that hot

flushes can occur even in the absence of pulsatile GnRH or LH release, implicating the involvement of upstream processes responsible for GnRH synthesis. The hypothalamic infundibular (arcuate) nucleus is a group of sex steroid-responsive neurons that co-express kisspeptin, dynorphin as well as NKB; the so-called 'KNDy' neurons[24]. Morphologic studies have shown that KNDy neurons from postmenopausal women are hypertrophied[25] and this hypertrophy is accompanied by elevated NKB, kisspeptin and Substance P gene expression; but not dynorphin[25-27]. These expression changes are a consequence of oestrogen withdrawal as they can be induced by oophorectomy and reversed by estrogen replacement in cynomolgus monkeys[28-30].

KNDy neurons, expressing NK_3 receptors, branch extensively within the arcuate nucleus and are also linked to the medial preoptic nucleus of the hypothalamus which has been identified as the control center for thermoregulation and which is responsive to both oestrogen and ambient temperature[31]. Support for the hypothesis that KNDy neurons have a thermoregulatory role and are involved in hot flush generation has been described in animal studies[32].

It is hypothesised that in the menopausal state the KNDy neurons are in a state of hyperactivation (consistent with the observed hypertrophy) which could disrupt baseline thermoregulation and trigger hot flushes[31]. Therefore, modulation of the KNDy neurons could be a viable therapeutic approach to controlling hot flushes. Clinical evidence supporting the hypothesis comes from a study in healthy women in which it was shown that intravenous infusion of NKB (endogenous ligand that binds the NK_3 receptor) acutely induces hot flushes in healthy women[33]. Furthermore, two recently completed proof of concept studies with NK_3 receptor antagonists have provided direct evidence that NK_3 receptor blockade can reduce the number and severity of hot flushes[34,35]. In a 12-week study in post-menopausal women, twice daily dosing with the NK_3 antagonist fezolinetant resulted in a sustained reduction in the frequency (-93% vs -54% for placebo) and severity (-70% vs -23%) of hot flushes (HF). The maximum improvements were apparent within 2 weeks of the start of treatment. In a 4-week study with the NK_3 antagonist MLE4901, the reduction in HF frequency was 78.4% compared to a reduction of 45.6% with placebo.

A role for SP and NK_1 receptors in the control of post-menopausal and other reproductive system disorders is supported by a body of evidence that indicates that the SP/NK_1 receptor system modulates KNDy neurons in unison with the NKB/NK_3 system. In adult male and/or ovariectomised female rodents, central infusion of specific NK_1 receptor agonists induces gonadotropin release[36,37]. In the mouse this response can be shown to be modulated by kisspeptin as knockout mice ($Kiss1r^{-/-}$) fail to respond to the NK_1 agonist[36]. Expression of the gene encoding for SP (Tacr1) can be inhibited by circulating estradiol in mice and ~50% of Kiss1 positive neurons also expressed the Tacr1 gene. In goats, however, although data suggests that SP (along with NKA and NKB) has a stimulatory effect on the GnRH pulse-generator and LH release, it only does so at high doses indicating that NKB in this species is the pivotal effector molecule[38]. Although there is evidence suggesting that SP does not play a key role in the non-human primate GnRH pulse generator[39], SP immunoreactivity frequently co-localises with kisspeptin and NKB in the human infundibular region[40]. Further, hypertrophy and increased gene expression of neurons containing SP messenger ribonucleic acid and SP+ immunoreactive nerves has been

demonstrated in humans in the hypothalami of postmenopausal women, with significantly more SP+ immunoreactive neurons than in age-matched men[41].

In addition to a central role in flushing, NK_1 receptors have been shown to be located in nerves associated with blood vessels in human skin and infusion of SP results in vasodilatation[42] and the same remote flushing of the face and neck characteristic of post-menopausal hot flushes [42,43].

As well as a role in improving vasomotor symptoms, NK_1 receptor antagonists have proven to be effective in Phase 2 studies in mood and primary insomnia[44,45,46,47] it is possible that the mood and sleep comorbidities associated with menopausal hot flushes could also be addressed with a molecule that has NK_1 receptor antagonist properties.

Taken together this evidence supports the hypothesis that the dual $NK_{1,3}$ receptor antagonism that NT-814 offers could be an effective treatment for vasomotor symptoms in the menopausal patient population as well as addressing important comorbidities.

5.3 Rationale for endpoints

There is no uniform clinical profile of hot flushes and variations are observed in terms of frequency, severity, duration, and course. Patients' perception of a hot flush and their impact on quality of life (QoL), sleep and psychosomatic reactions differ, and each is vulnerable to confounders, such as current mood, environmental factors or stress. It is, therefore, important to use a variety of assessment tools in clinical studies of hot flushes.

The primary efficacy assessment in this study will be based on subject reported assessments of hot flush frequency and severity. These will be recorded daily in an electronic diary (eDiary) throughout the study. Hot flush is a subjective experience for the subject and so its occurrence and intensity can only be assessed by the subject. The assessment will be based on the subject's recollection of the number and severity of hot flushes experienced twice daily; once in the evening on retiring to bed and once in the morning on waking. The morning assessment also enables the number of times the subject awoke at night to be recorded. These assessments are routinely employed in clinical studies in hot flush and their use is supported by recommendations from experts in the field[48] and regulatory authorities[49].

The impact on sleep will be measured using the Pittsburgh Sleep Quality index which is a validated questionnaire that evaluates a number of dimensions of sleep[50]. Insomnia, a common symptom of the menopause, will be assessed using the specific Insomnia Severity Index assessment. The broader impact of menopausal symptoms on subjects' health and well-being will be evaluated using the Menopause Specific Quality of Life Scale (MenQOL) which is a quality of life instrument designed specifically for use in the menopause and is used frequently in menopause clinical research[51]. In addition, the Hot Flash Related Daily Interference Scale (HFRDIS) will be used to evaluate interference with daily functioning. This is a validated instrument that has been shown to be sensitive to the effects of pharmacological interventions on menopause symptoms[52,53].The effect of NT-814 on symptoms of depression will be assessed

using version II of the Beck Depression Inventory. This is a widely used and validated self-reported tool for assessing both the psychological and physical symptoms of depression[54].

The PK exposure-response relationship will be established through the correlation of plasma NT-814 concentrations derived from sparse sampling, with appropriate clinical efficacy endpoints. Sparse sampling is a well-established method of collecting PK data that avoids the need for multiple blood draws and extended clinic visits.

The safety of NT-814 will be assessed using standard measures of safety appropriate for the stage of development (adverse events, routine safety laboratory tests, vital signs, electrocardiogram (ECG)). Although the effect is expected to be modest in the post-menopausal population, NK_3 receptor blockade may result in a reduction in circulating oestrogen concentrations and so safety assessments will, additionally, include an assessment of bone turnover. No adverse effects are anticipated but because NT-814 acts centrally in the nervous system its safety related to suicidal ideation will, in accordance with regulatory expectations, be assessed using the electronic version of the Columbia Suicide Severity Rating Score.

The assessment of safety for hormonal menopause therapies typically includes evaluation of the endometrium by trans-vaginal ultrasound (TVU) and, if indicated, endometrial biopsy. Since NK receptor blockade will, if anything, reduce hormonal drive to the endometrium these invasive evaluations are not warranted routinely for all subjects. However, any subject experiencing uterine bleeding (including spotting) during the study will undergo TVU and biopsy if indicated.

6. STUDY OBJECTIVE AND PURPOSE

6.1 Primary objectives

- To evaluate the efficacy of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, in reducing the frequency and severity of hot flushes.
- To assess the safety and tolerability of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, in subjects with post-menopausal symptoms.

6.2 Secondary objectives

- To evaluate the effect of once daily doses of 40 mg, 80 mg, 120 mg and 160 mg NT-814, compared with placebo, on mental well-being, quality of life and measures of sleep in subjects with post-menopausal symptoms.

- To evaluate the dose-response relationship of 40 mg, 80 mg, 120 mg and 160 mg NT-814 in reducing hot flush frequency and severity.

6.3 Pharmacokinetic objectives

- To evaluate the exposure-response relationship of NT-814 using population pharmacokinetics (PK) with sparse sampling.

6.4 Pharmacodynamic objectives

- Not applicable

7. SELECTION AND WITHRAWAL OF SUBJECTS

7.1 Subject numbers

Participants in the clinical investigation are referred to as "subjects".

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 (see Section 11).

The maximum number of subjects randomised into the study will be 300.

7.2 Inclusion Criteria

Subjects must meet all the inclusion criteria listed below.

1. Females aged 40 to 65 years, inclusive, at Screening Visit 1.

2. Able to understand and comply with the requirements of the study and give informed consent.

3. Postmenopausal, defined as: (i) at least 12 months of spontaneous amenorrhea, or (ii) at least 6 months of spontaneous amenorrhea with serum FSH levels > 40 mIU/ml and a serum oestradiol concentration of < 30 pg/mL, or (iii) at least 6 weeks postsurgical bilateral oophorectomy with or without hysterectomy[5],[6].

4. Body mass index between 18 and 38 kg/m^2, inclusive, at Screening Visit 2.

5. Negative urinary pregnancy test at Screening Visit 2.

6. In good general health, in the opinion of the investigator, based on the medical history, physical examination, 12-lead ECG, vital signs and clinical laboratory tests assessed at Screening Visit 2.

7. Subject has completed the paper diary for at least 6 days between Screening Visits 1 and 2 and has recorded an average of at least 8 moderate or severe hot flushes per day (including night-time) over the last 5 days that the paper diary was completed (assessed at Screening Visit 2).

[5] Durations are relative to screening visit 1

[6] Subjects who do not clearly fall into 1 of the 3 defined categories, but who are hormonally post-menopausal (serum FSH levels > 40 mIU/ml and a serum oestradiol concentration of < 30 pg/mL) and of an appropriate age, may also be considered for inclusion.

8. Subject has completed the electronic diary for at least 9 days between Screening Visit 2 and Day 1 and has recorded an average of at least 7 moderate or severe hot flushes per day (including night-time) over the last 7 days that the eDiary was completed (assessed at the Baseline Visit).

7.3 Exclusion Criteria

Subjects who meet one or more of the following exclusion criteria will not be enrolled.

1. Have used or unwilling to wash-out use of any of the following hormonal therapies for the periods stated prior to Screening visit 2:
    - ≥1 week for vaginal hormonal products (rings, creams, gels and including DHEA or analogues thereof)
    - ≥4 weeks for transdermal oestrogen alone or oestrogen/progestin products
    - ≥8 weeks for oral oestrogen (including selective oestrogen receptor modulators) and/or progestin therapy
    - ≥8 weeks for intrauterine progestin therapy
    - ≥3 months for progestin implants and oestrogen alone injectable drug therapy
    - ≥6 months for oestrogen pellet therapy or progestin injectable drug therapy
2. The use of non-hormonal prescription (eg paroxetine, other anti-depressants, alpha agonists [eg clonidine], methyldopa, gabapentin, pregabalin, medicinal cannabis) or over the counter/herbal treatments for treatment of menopausal symptoms is not allowed throughout the study. Subjects must have discontinued these drugs at least 28 days prior to Screening Visit 2. Subjects may, however, be permitted to continue to use these drugs if the dose has been stable for at least 4 weeks and they have been prescribed solely for the management of another disorder (e.g. neuropathic pain, depression).
3. Inability to comply with the use of prohibited medications as described below (See Appendix A for more details):
    i. Use of digoxin is not allowed from Screening Visit 2 until 1 week after the last dose of investigational medicinal product (IMP)
    ii. Use of known CYP3A4 substrates with a narrow therapeutic range is not allowed from Screening Visit 2 until 1 week after the last dose of IMP
    iii. Use of strong or moderate inhibitors of CYP3A4 is not allowed from Screening Visit 2 until 1 week after the last dose of IMP
    iv. Use of moderate or strong inducers of CYP3A4 is not allowed from Screening Visit 2 until Week 12
    v. Use of known P-glycoprotein inhibitors is not allowed from Screening Visit 2 until 1 week after the last dose of IMP
4. Any prior or ongoing history of clinically relevant drug or alcohol abuse within 12 months of Screening Visit 1.

5. Any clinically significant prior or ongoing history of arrhythmias, either determined through clinical history or on ECG evaluation.
6. Any clinically significant abnormal laboratory test result(s) measured at Screening Visit 2.
7. Any active ongoing condition that could have caused difficulty in interpreting vasomotor symptoms such as: infection that could have caused pyrexia, pheochromocytoma, hyperthyroidism, carcinoid syndrome, alcohol abuse.
8. Current history or previous (within the past 5 years) history of any malignancy (except basal and squamous cell skin tumours).
9. Uncontrolled hypertension[7].
10. A history of hyperthyroidism or hypothyroidism. Treated hypothyroidism with normal thyroid function test results at Screening Visit 2 and a stable (for ≥3 months before Screening Visit 2) dose of replacement therapy is acceptable.
11. Known hypersensitivity to NT-814 or any of the excipients in the formulation.
12. Concurrent (or within the 2 months prior to Screening Visit 1) participation in a clinical study with an investigational medicinal product.
13. Concurrent (or within the 1 month prior to Screening Visit 1) participation in an interventional clinical study.
14. Previous participation in a clinical study with NT-814.
15. Dependent on the investigator, the contract research organisation(s) or Sponsor for education or employment.
16. Any unexplained post-menopausal bleeding.
17. Abnormal findings on mammogram or subject has not undergone a mammogram within the guidelines recommended by applicable national authorities (eg United Kingdom National Health Service, United States Preventative Services Taskforce, Canadian Task Force on Preventative Healthcare)[8].

7.4 Withdrawal Criteria

Subjects will be informed that they have the right to withdraw from the study at any time without stating a reason and without prejudice to further treatment.

The Investigator may withdraw subjects from the study or may discontinue study treatment at any time.

[7] As a guide, a blood pressure of ≥140 mmHg (systolic) or ≥90 mmHg (diastolic) will usually require further evaluation. This may include a repeat after a period of rest/relaxation.
[8] Subjects who have not had a mammogram within the guidelines may have one performed as part of the screening process

Early withdrawal or early discontinuation from the IMP of any subject who has given informed consent to participate will be recorded including the reason for withdrawal. The primary reason will be selected from the following standard categories of early discontinuations or withdrawals:

- **Failed to meet randomisation criteria**
- **Adverse Event (AE)**: Clinical events occurred, or laboratory results are reported, that in the medical judgment of the investigator are grounds for discontinuation from participation in the best interests of the subject
- **Withdrawal of Consent**: The subject desired to withdraw from further participation in the study. The subject is not obliged to provide any reason for withdrawal of consent, but where a reason is given this will be recorded on the electronic Case Report Form (eCRF)
- **Protocol Violation**: The subject failed to adhere to the protocol requirements, and, in the opinion of the investigator, this failure to comply increased the risk to the subject to an unacceptable level
- **Lost to Follow-Up**: The subject stopped coming for visits and study personnel were unable to contact the subject.
- **Pregnancy**: Any subject becoming pregnant during the study will be withdrawn from dosing.
- **Other:** The subject was terminated for a reason other than those listed above

Subjects who withdraw or are withdrawn from participation in the study should attend an Early Termination visit at which the procedures scheduled for the final Follow Up visit will be undertaken. If the subject withdraws due to an AE, the event should be followed until resolution or care is transferred to the subject's usual physician.

Subjects who are withdrawn after randomisation will not be replaced.

In some circumstances, the Investigator may discontinue study treatment without the need to withdraw the subject from participation in the study. When subjects permanently discontinue IMP, effort should be made to continue following the subject through to the end of scheduled visits for safety and to obtain efficacy assessments that can be assigned to the originally randomised treatment.

7.5 Criteria for Stopping the Study

The Sponsor may terminate the study for safety or administrative reasons at any time.

The safety of NT-814 and the risk-benefit profile of individual doses and the study overall will be monitored by the study Data Review Committee (DRC). Full details, including specific criteria that may lead to study suspension or termination are described in the SWITCH-1 DRC Charter.

8. STUDY DESIGN

8.1 Co- Primary Efficacy Endpoints

There are four co-primary efficacy endpoints:

- Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Week 4
- Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Week 12
- Mean change from baseline in severity of moderate and severe hot flushes from baseline to Week 4
- Mean change from baseline in severity of moderate and severe hot flushes from baseline to Week 12

8.2 Secondary efficacy endpoints

The secondary efficacy endpoints include:

- Mean change from baseline in frequency of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in severity of moderate and severe hot flushes from baseline to Weeks 1, 2, 8 and 16
- Mean change from baseline in frequency of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in severity of all hot flushes from baseline to Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in the Hot Flush Score (frequency x severity) at Weeks 1, 2, 4, 8, 12 and 16
- Responder analyses: proportion of subjects with $\geq$50% and $\geq$80% reduction from baseline in hot flush frequency at Week 12
- Mean change from baseline in number of night time awakenings secondary to hot flush at Weeks 1, 2, 4, 8, 12 and 16
- Mean change from baseline in number of all night time awakenings 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

8.3 Pharmacokinetic endpoints

- Exposure-response modelling will be undertaken on a number of efficacy and safety endpoints on an exploratory basis.

8.4 Pharmacodynamic endpoints

- Not applicable

8.5 Safety endpoints

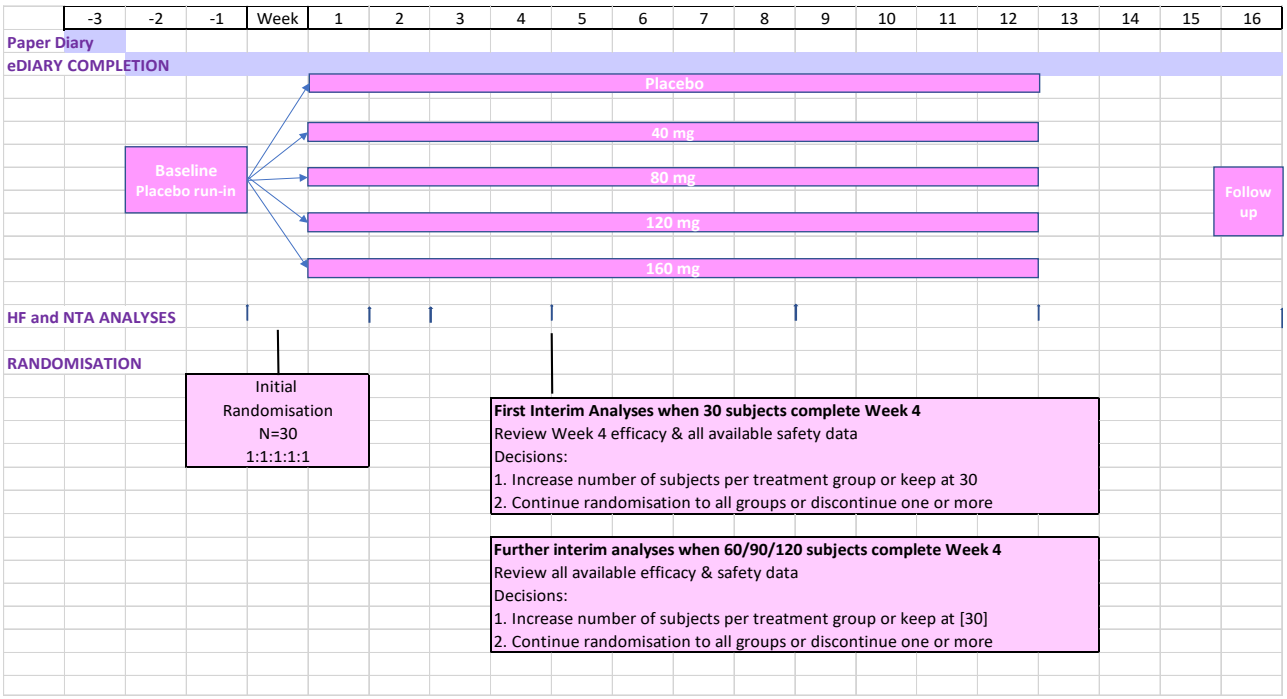
- Change from baseline in clinical laboratory assessments
  - At Weeks 2, 4, 12 and 16 for haematology and urinalysis
  - At Weeks 2, 4, 8, 12 and 16 for clinical chemistry
- Change from baseline in vital signs (blood pressure, pulse rate, temperature) at Weeks 2, 4, 8, 12 and 16
- Change from baseline 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 maximum absolute QTcF values by category at each visit
  - $\leq$450, >450 to $\leq$480, >480 to $\leq$500, >500 msec
- Proportion of subjects with maximum change from baseline in ECG QTcF values by category at Weeks 2, 4, 8, 12 and 16
  - $\leq$0, >0 to $\leq$30, >30 to $\leq$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

8.6 Study design

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

Subjects will be provided the Participant Information Sheet (PIS) and Informed Consent Form (ICF) for their prior review. Informed consent to participation in the study will be obtained for all subjects before any study related procedures are performed.

Figure 1: Study schematic



Subjects will be recruited initially into five parallel groups: 40 mg once per day NT-814, 80 mg once per day NT-814, 120 mg once per day NT-814, 160 mg once per day NT-814 and placebo once per day. See Figure 1 for study design schematic. After 30 subjects have been randomised and completed all assessments through the Week 4 visit the emerging efficacy and safety data will be reviewed by the DRC and the first determination made as to whether randomization to all dose groups should be continued. Further interim reviews may be conducted after 60, 90, 120 and 150 subjects have completed the Week-4 assessments. Once a subject has been allocated to a dose/treatment group they will continue to receive the same dose/treatment throughout the study. Subjects will continue to be allocated to placebo treatment throughout the study. Further details of the DRC reviews are provided in Section 11.2 and in the DRC Charter.

8.6.1 Justification for Study Design

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 which broadly encompasses the equivalent exposures to those evaluated in the RELENT-1 study that were associated with efficacy (See Section 5.1.3).

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. The interim reviews also enable the sample size (per treatment group) to be assessed on an ongoing basis and to be increased if the estimates of treatment effect and variability used to derive the sample size originally prove to be incorrect. These adaptive features ensure that the study has the greatest probability of success while making efficient use of the trial population.

The study will be placebo controlled to minimise assessment bias. Placebo is considered an acceptable reference group for studies in PM VMS as the disorder, whilst uncomfortable and disruptive to the affected individual, is not life-threatening or limiting, and studies are typically of limited duration. Experience from multiple studies in women with VMS has shown that a placebo response of up to 60% can be expected[48] and so even subjects allocated to placebo should expect to receive some benefit from study participation. The use of placebo is also recommended by both the United States FDA and EU CHMP disease-specific guidelines[49,55].

Although the interim analyses used to make the adaptive randomisation decisions will be reviewed by an unblinded Data Review Committee (DRC), the Sponsor and CRO study operational team will remain fully blinded throughout the study. The subjects and investigative site staff will also be blinded throughout the randomised phase of the trial and thus the study will be triple blind to reduce assessment bias. All subjects will receive placebo on a single blind basis during the last 2-weeks of the baseline period to enable baseline assessments to be recorded under similar conditions to those used during the randomised phase of the study. This run-in period, which has been used in similar trials previously[56], also allows participants to familiarise themselves with the requirements and systems used in the study before the start of the randomised treatment phase.

The study will comprise a 2-week baseline period followed by a 12-week treatment period. The 2-week baseline period allows subjects one week to become familiar with eDiary completion before the baseline efficacy assessments are made during the second week. Studies with other NK antagonists have demonstrated that maximum efficacy is achieved within a matter of weeks[34,35], and, in the case of fezolinetant, is sustained for at least 12-weeks. The disease-specific guidelines from both the FDA and CHMP recommend studies of 12-weeks duration to demonstrate maintenance of effect[49,55].

8.6.2 Duration of Subject Participation

Subjects will participate in the study for a total of approximately 19 weeks, comprising a formal screening and 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.

8.6.3 Screening Visit 1

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.

The visit will include the following assessments to determine the subject's initial eligibility for inclusion into the study:

- Recording of the subject's medical history, and prior and concomitant medication use, in particular, hormonal therapies and drugs used to manage menopausal symptoms

Subjects will be provided with, and trained on the use of, a paper diary to be completed once a day for 7 days. Each evening (before retiring to bed), subjects will record in the diary their recall of the total number of hot flushes of each severity experienced over the previous 24 hours.

The minimum time between Screening Visits 1 and 2 is 6 days, to enable at least 6 days of paper diary data to be completed. If necessary, the time period between Screening Visits 1 and 2 can be extended to enable prohibited concomitant medications to be washed-out.

8.6.4 Screening Visit 2

Subjects will return to the clinic for Screening Visit 2, during which the investigator or designee will review the paper diary to determine if the subject can continue in the study. To continue, subjects must have completed the paper diary for at least 6 days and have recorded an average of at least 8 moderate or severe hot flushes per day over the last 5 days that the paper diary was completed.

Subjects who **do not** record an average of at least 8 moderate or severe hot flushes per day (including night-time) over the last 5 days that the paper diary was completed, or who are not compliant with diary completion will be Screen Failures and will not be able to continue into the study. Limited data will be collected for subjects who withdraw from the study after providing informed consent but prior to randomisation onto the study (Screen Failures).

If the diary compliance and hot flush criteria **are** satisfied, the following assessments will be performed to further determine the subject's eligibility for inclusion into the study:

- Review of other inclusion/exclusion criteria
- Full physical examination, 12-lead ECG, vital signs
- Blood sampling for haematology and clinical chemistry, and urine sample collection for urinalysis and urine pregnancy test
- Review for adverse events and changes in concomitant medications

Subjects who continue to satisfy selection criteria will be provided with, and trained on the use of, an eDiary to be completed twice a day, starting the evening of Screening Visit 2 and continuing until they return to the site for the Baseline visit. Each evening, subjects will record in the diary the total number of hot flushes of each severity experienced that day since waking. Each morning upon waking, subjects will record the number of times they woke up in the night and the total number of hot flushes of each severity experienced during the night.

At Screening Visit 2, subjects will commence placebo treatment. 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. Subjects will be provided with a medication kit containing sufficient placebo capsules for a 2-week period (plus 2 days overage).

This placebo-run in period will be single-blind in design, in that whilst the investigator and clinic staff will know that placebo is being provided, subjects must remain blinded, and should be merely be informed that they are being given study medication.

8.6.5 Baseline (Day 1)

Subjects will return to the clinic for the Baseline visit at which compliance with eDiary completion and hot flush data will be reviewed. Subjects who comply with all other selection criteria, who have completed the diary for at least 9 days, and who have recorded an average of at least 7 moderate or severe hot flushes per day (including night-time) over the last 7 days that the eDiary was completed may enter the 12-week double-blind portion of the study and be randomised to one of 40 mg, 80 mg, 120 mg or 160 mg NT-814 or placebo treatments. The mean number of hot flushes will be calculated programmatically by the eDiary vendor and the outcome made available to the site.

Subjects who **do not** record an average of at least 7 moderate or severe hot flushes per day (including night-time) over the last 7 days that the eDiary was completed, or who are not compliant with eDiary completion will be deemed Screen Failures and will not be able to continue into the study. Limited data will be collected for subjects who withdraw from the study after providing informed consent but prior to randomisation onto the study (Screen Failures).

The Baseline visit will be Day 1. Prior to dosing on Day 1, the following assessments will be completed:

- Review of inclusion/exclusion criteria, including eDiary data
- Symptom directed physical examination
- Vital signs (after resting for ≥ 5 mins, see Section 10.3.8)
- 12-lead ECG
- Blood samples for clinical chemistry, haematology and plasma bone turnover markers
- Urine sample for urinalysis
- Subjects will be asked to complete the MenQoL-I, HFRDIS, Pittsburg Sleep Quality Index, Insomnia Severity Index, electronic Columbia Suicide Severity Rating Scale (eC-SSRS) and the Beck Depression Inventory II.
- Review for adverse events and changes in concomitant medications

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).

Subjects will be given their first dose of IMP in the clinic and take all subsequent doses at home once a day in the evening (starting the following day) for a total of 12 weeks. At the baseline visit, subjects will be provided with IMP sufficient for the next 4 weeks.

Subjects will continue to complete the eDiary twice a day throughout the study until the Final Follow-up visit at Week 16.

8.6.6 Interim visits (Weeks 2, 4, 8, and 12)

During the double-blind treatment phase, subjects will return to the clinic for interim assessments at Weeks 2, 4, 8 and 12. Weeks 2 and 4 should be scheduled within ±3 days of the target visit day and Weeks 8 and 12 within ±4 days of the target visit day. The visit at Week 12 should be conducted at a similar time of day to the baseline visit to reduce diurnal variation in the plasma bone turnover markers.

At the start of each visit a blood sample will be taken for determination of plasma levels of NT-814. At the same time, if required per the Study Schedule, blood samples for safety laboratory assessments and plasma bone turnover markers will be taken. Subjects will be asked to provide the time they took their dose of study medication on the previous day and this will be recorded on the eCRF.

Study medication compliance will be reviewed at all visits and a new supply of study medication (sufficient for the next 4 weeks) will be provided at Weeks 4 and 8. Diary completion compliance will also be reviewed, and training/re-training provided as necessary.

At each visit the following assessments will be completed:

- Vital signs (after resting for ≥ 5 mins, see Section 10.3.8)
- 12-lead ECG
- Blood samples for hematology (Weeks 2, 4 and 12), clinical chemistry (Weeks 2, 4, 8 and 12), pharmacokinetic analysis (Weeks 2, 4, 8 and 12) and plasma bone turnover markers (Week 12 only)
- Urine samples for urinalysis (weeks 2, 4 and 12)
- Subjects will be asked to complete the MenQoL-I, Pittsburgh Sleep Quality Index and Insomnia Severity Index (Weeks 4, 8 and 12), the HFRDIS and Beck Depression Inventory II Weeks 2, 4, 8 and 12) and eC-SSRS (Weeks 4 and 12)
- Review of eDiary compliance
- Review for adverse events and changes in concomitant medications

At Week 12, subjects may be re-started on any prohibited medications that are permitted after the end of treatment (see Appendix A).

8.6.7 Final Follow-up / Early Termination visit

Subjects will return to the clinic for the final study visit 4-weeks (±5 days) after the end of treatment. The visit should be conducted at a similar time of day to the baseline visit to reduce diurnal variation in the plasma bone turnover markers.

The same procedures will also be undertaken if a subject withdraws from the study early.

The eDiary will be collected and the following assessments completed:

- Symptom directed physical examination
- Vital signs (after resting for ≥ 5 mins per Section 10.3.8)
- 12-lead ECG
- Blood samples for clinical chemistry and hematology and plasma bone turnover markers
- Urine sample for urinalysis
- Subjects will be asked to complete the MenQoL, HFRDIS, Pittsburg Sleep Quality Index, Insomnia Severity Index, eC-SSRS and the Beck Depression Inventory II
- Review for adverse events and changes in concomitant medications

Subjects may be re-started on any prohibited medications and menopausal symptom management medications at this time.

9. INVESTIGATIONAL PRODUCT AND ADMINISTRATION

9.1 Selection of doses in the study

Subjects will initially be randomised to one of four NT-814 dose regimens and placebo. The use of four active doses and placebo will enable an adequate range of doses and exposures and several dose-response scenarios to be evaluated.

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.

Two reference points have been used to set the doses in the current study. The first is the efficacy findings in the RELENT-1 proof of concept study. Ideally, the pattern of response with respect to

dose will be replicated in the current study but the treatment groups in RELENT-1 were small and interpretation of the findings was complicated by the variable pharmacokinetics of the hard gel capsule formulation used in that study. A 150 mg dose of NT-814 was found to be maximally effective with no greater efficacy observed at the higher dose of 300 mg. A dose of 100 mg was effective in some endpoints but was clearly sub-optimal in comparison to the 150 mg dose. A dose of 50 mg was ineffective. The RELENT-1 study sets the target exposures for future studies.

Table 9-1 Summary of doses and exposures in the RELENT-1 Proof of Concept Study

Dose	Mean $AUC_{0\text{-last,ss}}$ (ng*hr/mL)	Mean $AUC_{0\text{-24,ss}}$ (ng*hr/mL)	Effectiveness
50 mg	2,341	2,341	Ineffective
100 mg	3,503	3,542	Sub-optimal
150 mg	9,021	5,165	Highly effective
300 mg	25,874	14,822	No additional benefit over 150 mg

$AUC_{0\text{-last,ss}}$ = Area under the plasma concentration-time curve from 0 to last measurable time-point at steady state

Since the hard gel capsule formulation used in RELENT-1 had unacceptable variability in kinetics, an optimised soft-gel capsule formulation has been developed for use in this and future studies. A relative bioavailability (RBA) study has been conducted which confirmed an acceptable level of variability and showed that, compared to the exposures seen in RELENT-1 with the hard gel capsule formulation, the soft-gel results in approximately twice the exposure on a "per mg" basis (Table 9-2). These data are the second reference point for setting the doses in the current study.

Table 9-2 Comparison of Exposures with single doses of the soft and hard gel capsule formulations of NT-814

Geometric mean	C_{max} ng/ml	$AUC_{0\text{-last}}$ ng*hr/ml	$AUC_{0\text{-24}}$ ng*hr/ml
25 mg soft gel (RBA Study)	332	1,010	786
150 mg hard gel (RELENT-1 Study)	911	3,012	3,013

$AUC_{0\text{-24}}$ = Area under the plasma concentration-time curve from 0 to 24 hours, $AUC_{0\text{-last}}$ = Area under the plasma concentration-time curve from time 0 to the last measurable concentration, C_{max} = maximum plasms concentration

A 40 mg soft-gel capsule has been developed and so all doses in the current study will be a multiple of 40 mg. The exposure after a single 150 mg dose in RELENT-1 is approximately 3-fold higher than the exposure after a single 25 mg dose of the soft gel capsule. Thus, an equivalent dose in the soft gel capsule is ~80 mg. In all previous clinical evaluations the kinetics of NT-814 have been found to be either linear or very close to linear with respect to dose and so linearity is also assumed when setting the doses for the current study. Assuming a 'pivot' dose of 80 mg, dose-linearity across the dose range and an accumulation ratio of 1.7 (to convert single dose data to steady state conditions), the estimated median AUCs at steady state at the four doses planned will be as shown in Table 9-3.

Table 9-3 Estimated NT-814 exposures at steady state

Dose	$AUC_{0\text{-last}}$ at Steady State		
	Lower bound 95% CI	Geometric Mean	Upper Bound 95% CI
40 mg	2,390	2,740	3,150
80 mg	4,780	5,500	6,310
120 mg	6,760	8,240	9,460
160 mg	9,580	10,990	13,350

$AUC_{0\text{-}24}$ = Area under the plasma concentration-time curve from 0 to 24 hours, CI = confidence interval

These four doses enable a range of exposures to be evaluated that encompass the full range of effectiveness responses observed in RELENT-1. The mean values at the mid doses (80 and 120 mg) are around the point that exposure-response modelling predicts maximum efficacy and the low and high doses are, respectively, expected to enable the rising and plateau parts of the dose-response curve to be demonstrated.

The upper bound of the confidence interval for predicted exposure at the high dose is below the mean exposure for the high dose in the RELENT-1 study (~15,000 ng*hr/ml) and is substantially below the individual highest exposures (up to ~80,000 ng*hr/ml) that were well tolerated in that study. It is also below the mean (~14,000 ng*hr/ml) and maximum (~42,000 ng*h/mL) values observed in healthy males after repeat dosing for 4 weeks that were also well tolerated with no safety concerns.

Thus, the administration of these four dose levels together with sparse PK sampling in this study will allow a robust evaluation of both the dose-response and PK exposure-response, from which it will be possible to select an appropriate dose to progress into the pivotal Phase 3 registration studies.

9.2 Investigational medicinal product

The IMP will consist of NT-814 40 mg capsules or matching placebo. The capsules are oblong, soft-gelatin capsules and each subject will take 4 capsules orally once-daily. The number of active and placebo capsules that each subject will take will be dependent on the dose to which they have randomized:

Dose	IMP Supply
Placebo	4 x placebo capsules
40 mg NT-814	1 x NT-814 capsule, 3 x placebo capsules
80 mg NT-814	2 x NT-814 capsules, 2 x placebo capsules
120 mg NT-814	3 x NT-814 capsules, 1 x placebo capsule
160 mg NT-814	4 x NT-814 capsules

Inactive ingredients in both the NT-814 capsules and the placebo capsules are Capmul MCM EP, Labrasol ALF, Tween 80, Peceol and Vitamin E.

The physical, chemical, pharmaceutical formulation properties and characteristics of the IMP are described in the IB.

All capsules will be blistered and packaged into weekly cards, with each weekly card containing capsules for a 7-day week plus an additional day for overage. These weekly cards will be further packaged into cardboard cartons:

- For the single-blind placebo run-in period, a carton will be dispensed containing 2 weekly placebo cards thereby providing sufficient capsules for the 2-week period between Screening Visit 2 and the Baseline visit (plus 2 days overage to allow for delays in clinic visit scheduling).

- For the post-randomisation double-blind treatment phase, each carton will contain 4 weekly cards thereby providing sufficient capsules for a 4-week dosing period (plus 4 days overage to allow for delays in clinic visit scheduling).

The IMP cards and cartons will be labelled in accordance all applicable local regulatory requirements. Multi-language labels may be applied to the IMP.

All IMP must be stored in a secure area with access limited to the Investigator and authorized site staff. The IMP is to be shipped and stored at controlled room temperature and be protected from light.

Only authorized investigational site study staff members are to dispense the IMP.

9.3 Allocation to Treatment

The 2-week placebo-run in period will be conducted in a single-blind manner, with the subjects blinded to the treatment allocated.

The post-randomisation 12-week treatment phase of the study will be conducted in a double-blind manner, with the subjects, Investigators and Sponsor all blinded to the treatment allocated. Both NT-814 and placebo will be presented as soft-gel capsules, identical in size and shape.

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 may be changed (see Section 11.2). 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 IMP each subject will receive will be allocated by an IWRS tool provided by Pharm Olam on behalf of the Sponsor.

9.4 Study Treatment and Administration

IMP will be taken once daily both during the 2-week placebo run-in phase and the 12-week double-blind treatment phase.

For the placebo run-in, the first dose will be taken in the clinic at Screening Visit 2 and the last dose will be taken the evening before the Baseline visit.

For the double-blind treatment phase, the first dose will be taken in the clinic at the Baseline visit (Day 1) and the last dose will be taken the evening before the Week 12 visit. IMP doses are fixed and will not be adjusted for individual subjects during the study.

Subjects will be required to take 4 capsules of IMP once-daily in the evening before bedtime.

At Screening Visit 2, Baseline, Week 4 and Week 8 subjects will be dispensed sufficient IMP for daily dosing at home until the next visit, with overage included to allow for visit windows.

At Screening Visit 2, subjects will be dispensed sufficient supply to cover the period to the Baseline visit.

At the Baseline visit subjects will be dispensed sufficient supply to cover the period to Week 4. An interim check of compliance will be made at the Week 2 visit (subjects will be instructed to bring their IMP carton with them to this visit).

At the Week 4 visit subjects will be dispensed sufficient supply of IMP to cover the period from Week 4 to Week 8.

At the Week 8 subjects will be dispensed sufficient supply of IMP to cover the period from Week 8 to Week 12.

On the days of clinic visits (Baseline, Weeks 2, 4, 8, and 12) subjects will be asked to provide the time that they took their dose the previous evening and this will be documented in the eCRF.

All doses are to be taken at approximately the same time each evening. If a subject forgets to take a dose in the evening then the dose can be taken at any time up until 9 a.m. the following day. After this time, the dose should not be taken and will be considered a missed dose.

Subjects will return any unused medication at each clinic visit.

9.4.1 Dose Interruption

In the event that a subject experiences an AE which the investigator believes is treatment related and which the subject finds intolerable, a break in dosing of up to one week is permitted. If, on reintroduction of the study medication, the adverse event recurs and remains intolerable the study drug will be withdrawn altogether. A break in dosing will not result in extension of the overall dosing period.

9.5 Treatment Accountability and Compliance Checks

In accordance with regulatory requirements, the Investigator or designated site staff must document the amount of IMP dispensed and/or administered to study subjects, the amount

returned by study subjects, and the amount received from and returned to the Sponsor (or representative) when applicable. IMP accountability records must be maintained throughout the course of the study. The accountability unit for this study is a capsule. Discrepancies are to be reconciled or resolved. Procedures for final disposition of unused IMP will be provided in the appropriate Study Manual.

The Investigator has overall responsibility in ensuring that the study treatment is received and managed at the study site in accordance with Good Clinical Practice (GCP).

Limited responsibility may be delegated to a nominated study site representative and this delegation must be documented. Study treatment will be dispensed by a nominated person at each study site.

The Sponsor will be permitted upon request to audit supplies, storage and dispensing records and procedures.

Every effort should be made to encourage subject compliance with the dosage regimen as per the protocol. All subjects will be instructed to return their medication carton with any unused capsules at each visit. A record of the capsules dispensed, taken, and returned will be recorded at each visit and compliance calculated.

9.6 Treatment Blinding Code

Study investigators 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 possible, the CRO medical monitor or Sponsor medical expert should be consulted prior to the blind being broken. If this is not possible, the CRO medical monitor or Sponsor medical expert must be notified as soon as possible after a code break was performed.

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, in order to provide the Regulators the knowledge of the event and the causal agent. A blinded copy of the report will be provided to the investigators and the relevant IRB/IEC.

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.

9.7 Permitted Concomitant Medications

All concomitant medications taken during the study will be recorded in the eCRF with indication, dose information, and dates of administration.

Any medication that is not specifically prohibited is allowed.

9.8 Prohibited Concomitant Medication

A list of prohibited medications is provided in Appendix A.

The use of hormonal therapies is prohibited prior to Screening Visit 2 for the periods stated in exclusion criterion 1, and throughout out the study.

The use of non-hormonal prescription (eg paroxetine, other anti-depressants, alpha agonists [eg clonidine], methyldopa, gabapentin, pregabalin, medicinal cannabis) or over the counter/herbal treatments for treatment of menopausal symptoms is not allowed throughout the study. Subjects must have discontinued these drugs at least 28 days prior to Screening Visit 2. Subjects may, however, be permitted to continue to use these drugs if the dose has been stable for at least 4 weeks and they have been prescribed solely for the management of another disorder (e.g. neuropathic pain, depression).

The following medications are also excluded from Screening Visit 2 until the last dose of IMP (unless stated otherwise).

- Known CYP3A4 substrates with a narrow therapeutic range
- Strong or moderate inhibitors of CYP3A4
- Strong or moderate inducers of CYP3A4 (until Week 12 only)
- Known P-glycoprotein inhibitors
- Digoxin

9.9 Other Study Restrictions

Mild somnolence has been identified as a possible adverse reaction to NT-814. Subjects will be instructed neither to drive nor operate machinery if they experience somnolence.

10. STUDY SCHEDULE

Table 10-4: Schedule of Events

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

10.1 Volume of Blood Sampling

Approximate blood sample volumes are detailed below. The volumes for individual samples may vary accordingly to laboratory requirements and the number of samples taken may increase due to unscheduled visits, lost or damaged samples where additional blood draws are needed, however, the total volume is not expected to exceed 150 mL per subject.

Clinical chemistry	5 mL
Hematology	3 mL
PK	6 mL
Plasma bone turn-over markers	10 mL

10.2 Efficacy assessments

10.2.1 Hot Flush Electronic Diary

Subjects will be provided with an eDiary to document the number of hot flushes experienced, the severity of each hot flush using a scale of 1 to 3 (mild, moderate or severe) and the number of night-time awakenings (Appendix C). Subjects will complete the eDiary twice a day, starting the evening of Screening Visit 2 and continuing throughout the study until the final Follow up visit. Each evening, subjects will record the total number of hot flushes of each severity experienced that day since waking. Each morning upon waking, subjects will record the number of times they woke up in the night and the total number of hot flushes of each severity experienced during the night. The severity of each hot flush will be graded by the subject as follows:

Mild: Sensation of heat without sweating

Moderate: Sensation of heat with sweating, but able to continue activity.

Severe: Sensation of heat with sweating, causing cessation (stopping) of activity

A hot flush at night which awakens the subject should be classed as severe. Once awake, subjects should record any further hot flush episodes according to their severity.

10.2.2 MenQoL-I (1 month recall version)

The MenQoL-I (Menopause-specific Quality-of-Life Questionnaire Intervention Version, Appendix C) 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.

Subjects will complete the MenQoL-I questionnaire at each clinic visit except the two Screening visits and Week 2.

10.2.3 Hot Flash Related Daily Interference Scale (HFRDIS)

The HFRDIS (Appendix C) 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).

Subjects will complete the HFRDIS at each clinic visit except the two Screening visits.

10.2.4 Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI, Appendix C) 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. The sum of scores for these seven components yields one global score.

Subjects will complete the PSQI at each clinic visit except the two Screening visits and Week 2.

10.2.5 Insomnia Severity Index

The Insomnia Severity Index (ISI, Appendix C) 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.

Subjects will complete the ISI at each clinic visit except the two Screening visits and Week 2.

10.2.6 Beck Depression Inventory II

The Beck Depression Inventory II (BDI-II, Appendix C) 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. 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.

Subjects will complete the BDI-II at each clinic visit except the two Screening visits.

10.3 Safety Assessments

10.3.1 Medical history and concomitant diseases

A review of medical and medication history will be performed at the Screening visits to confirm subject eligibility. This must be confirmed by medically qualified investigator or sub-investigator. All medically and clinical relevant information must be recorded, regardless of the time since the date of diagnosis.

History should include (but is not limited to):

- All current and past medications taken during the six months before the Screening Visit
- Relevant history of menopausal, respiratory, cardiovascular, renal, gastro-intestinal, hepatic, endocrine, hematological, neurological, psychiatric and any other diseases

The menopausal symptom history does not need to be recorded on the general medical history page of the eCRF. However, surgeries related to menopause (e.g. bilateral oophorectomy, hysterectomy) do need to be recorded on the medical history page.

10.3.2 Physical examinations

A full physical examination will be conducted at Screening Visit 2 and a symptom directed physical examination at the Baseline and Week 16 visits. This will be completed by a physician or an appropriately qualified delegate.

A full 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
- Neurological

Any abnormalities that are identified at Screening Visit 2 will be documented on the medical history eCRF page. Any change (including new and worsening findings) between Screening Visit 2 and the final study visit should be recorded as an AE on the AE eCRF page and in the subject's medical record.

If an improvement/resolution of a physical examination finding documented in the subject's medical history occurs during the study, it should be recorded in the source document. If there is resolution of a physical examination finding previously noted as an AE, then the event resolution and stop date should be recorded on the AE eCRF page and documented in the subject's notes.

10.3.3 12-Lead ECGs

Resting 12-lead ECG data will be captured at all study visits except Screening Visit 1.

All ECGs will be performed after the patient has rested for five minutes in a semi-recumbent position. The same model of ECG recorder should be used throughout the study for any given subject wherever possible.

All ECG reports must be reviewed, signed and dated by the Investigator or delegated physician. Reports will then be filed with the subject's medical record. The Investigator will comment on all abnormal findings and determine whether they are clinically significant. These assessments will be recorded in the eCRF and clinically significant findings must also be reported as AEs in the eCRF.

10.3.4 Electronic Columbia Suicide Severity Rating Scale

The Columbia Suicide Severity Rating Scale, or C-SSRS (Appendix C), 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." The version used will be the electronic C-SSRS (eC-SSRS), which is a subject-reported version of the scale. The C-SSRS has been found to be reliable and valid in the identification of suicide risk in several research studies and is a standard tool used in clinical development studies with centrally acting investigational drugs, regardless of indication.

Subjects will complete the eC-SSRS at Baseline, Weeks 4, 12, and 16.

10.3.5 Clinical chemistry

Blood for clinical chemistry assessments will be collected as indicated in the Study Schedule (Table 10-1) and sent to the central laboratory for analysis. The following clinical chemistry parameters will be assessed: 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 and haemoglobin A1c (HbA1c). Subjects do not need to fast before clinical chemistry samples are taken.

All clinical chemistry test reports must be reviewed, signed and dated by the Investigator or delegated physician. Reports will then be filed with the subject's medical record. The Investigator will comment on all abnormal results and determine whether they are clinically significant. Clinically significant findings must also be reported as AEs in the eCRF.

10.3.6 Haematology

Blood for haematology assessments will be collected as indicated in in the Study Schedule (Table 10-1) and sent to the central laboratory for analysis. The following haematology parameters will be assessed: red blood cell (RBC) count, white blood cell (WBC) count, haematocrit, haemoglobin, MCV, platelet count and WBC differentials.

All haematology test reports must be reviewed, signed and dated by the Investigator or delegated physician. Reports will then be filed with the subject's medical record. The Investigator will comment on all abnormal results and determine whether they are clinically significant. Clinically significant findings must also be reported as AEs in the eCRF.

10.3.7 Urinalysis

Urine for urinalysis assessments will be collected as indicated in the Study Schedule in (Table 10-1) and sent to the central laboratory for analysis. Urinalysis will include glucose, bilirubin, ketones, specific gravity, blood, pH, protein, urobilinogen, nitrites, leucocyte esterase, sedimentation.

All urinalysis test reports must be reviewed, signed and dated by the Investigator or delegated physician. Reports will then be filed with the subject's medical record. The Investigator will comment on all abnormal results and determine whether they are clinically significant. Clinically significant findings must also be reported as AEs in the eCRF.

10.3.8 Vital signs

Vital signs will be measured at all visits except Screening Visit 1:

- 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.

At Screening Visit 2 only, blood pressure measurements may be repeated a maximum of twice (three measurements in total) if the first values are high and suggestive of uncontrolled hypertension. The repeat measurements should be taken after at least 5 minutes of rest in a quiet environment.

All measurements are to be recorded on the Vital Signs eCRF.

All vital signs must be reviewed by the Investigator or delegated physician. The Investigator will comment on all abnormal results and determine whether they are clinically significant. These assessments will be recorded in the eCRF and clinically significant findings must also be reported as AEs in the eCRF.

10.3.9 Bone Turnover Markers

Blood for assessment of bone turnover markers will be collected at Baseline, Week 12 and Week 16 and sent to the central laboratory for analysis. The serum concentration of the following markers will be assessed: bone-specific alkaline phosphatase (BALP) and procollagen type 1 N-terminal pro-peptide (P1NP).

Baseline and Week 12 and 16 visits should be conducted at similar times of day to reduce diurnal variation in these markers.

10.3.10 Pregnancy Testing, Contraception and Pregnancy

Pregnancy Testing

A dipstick urine pregnancy test will be performed for all subjects at Screening Visit 2 and must be negative for the subject to remain eligible.

Contraception

There are no contraception requirements given that all subjects will be post-menopausal.

Pregnancy

Not applicable for this study.

10.4 Pharmacokinetics

Blood samples for analysis of plasma NT-814 concentrations will be collected at Weeks 2, 4, 8, and 12. The actual date and time of each blood sample collection will be recorded, together with the date and time of the most recent dose of IMP.

Approximately 6 mL of blood will be collected at each time point. Details of PK blood sample collection (including volume to be collected), processing, storage and shipping procedures will be provided in the Study Manual.

10.5 Adverse events

10.5.1 Definitions

Adverse Event (AE)

Any untoward medical occurrence in a subject or clinical trial subject administered an IMP and which does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding, for example), symptom, or disease temporally associated with the use of an IMP, whether or not considered related to the IMP.

As the full medical history and physical evaluation does not take place until Screening Visit 2, the adverse event reporting period starts after this visit is complete.

Adverse Drug Reaction (ADR)

All AEs considered to be untoward and unintended responses to an IMP related to any dose should be considered ADRs. The definition covers also medication errors and uses outside what is foreseen in the protocol, including misuse and abuse of the product. The phrase "responses to an IMP" means that a causal relationship between an IMP and an AE is at least a reasonable possibility, i.e., the relationship cannot be ruled out.

Serious Adverse Event (SAE)

An adverse event that:

- Results in death
- Is life-threatening (i.e. the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it was more severe)
- Requires in-patient hospitalization or prolongation of existing hospitalization (see explanation below)
- Results in persistent or significant disability/incapacity
- Is a congenital anomaly/birth defect
- Is considered to be an important medical event

Based upon medical and scientific judgment, important medical events that may not be immediately life-threatening, or result in death or hospitalization, but may jeopardize the subject and/or[9] may require intervention to prevent one of the other outcomes listed in the definition above may be considered a SAE.

Hospitalizations are defined as initial or prolonged admissions that include an overnight stay. Hospitalization or prolonged hospitalization for technical, practical or social reasons, in the absence of an adverse event is not an SAE and neither is attendance at an Emergency Room / Accident and Emergency Department that takes place in the evening or night and does not result in formal admission to the hospital.

Unexpected Adverse Drug Reaction

An adverse reaction, the nature or severity of which is not consistent with applicable product reference safety information (Section 6.5 of the Investigator Brochure).

[9] The definition varies between regulatory jurisdictions

Reports which add significant information on the specificity, increase of occurrence, or severity of a known, already documented serious adverse reaction constitute unexpected events.

Suspected Unexpected Serious Adverse Reaction (SUSAR)

A serious adverse event that is suspected to be related to the administered medicinal product and the nature or severity of which is not consistent with the reference safety information.

Adverse Events of Special Interest

NT-814 is a non-hormonal therapy and so is not expected to have any adverse effect on the endometrium. However, post-menopausal bleeding will be monitored as an AE of special interest and investigators should report any such events to the Sponsor in an expedited manner. This should be done by completing an AE of Special Interest form within 5 working days of awareness of the event and submitting the form to the medical monitor by email. Note, the AESI form should NOT be sent to Emas.

Additionally, any subjects experiencing post-menopausal bleeding after randomisation should undergo a transvaginal ultrasound with subsequent investigation and management (including endometrial biopsy, if indicated) according to the investigator's clinical judgement and usual practice (unexplained post-menopausal bleeding *prior* to randomisation will exclude the subject from the study).

10.5.2 Assessment of Severity

The severity (intensity) of each adverse event will be classified as:

- **Mild** Awareness of sign or symptom, but easily tolerated
- **Moderate** Sign or symptom causes discomfort, but does not interfere with normal activities
- **Severe** Sign or symptom of sufficient intensity to interfere with normal activities

10.5.3 Assessment of Causality

A determination will be made of the relationship (if any) between an AE and the study drug. A causal relationship is present if a determination is made that there is a reasonable possibility that the AE may have been caused by the drug or study procedure. AEs will be classified as either related or unrelated.

10.5.4 Adverse Event Reporting

Adverse events may be volunteered spontaneously by the subject, or discovered as a result of general, non-leading questioning. As the full medical history and physical evaluation does not take place until Screening Visit 2, the adverse event reporting period starts after this visit is

complete. Adverse events occurring from the time of Screening Visit 2 up to the final study visit will be recorded. All AEs should be recorded in the eCRF and in the subjects' source notes.

Any SAE occurring from Screening Visit 2 until 28 days after the last dose of IMP must be reported immediately (within 24 hours of the investigator becoming aware of it) and recorded on the SAE Form. Detailed instructions for the reporting of SAEs can be found in the Study Manual. All subjects with a SAE must be followed up and the outcomes reported. The investigator must supply the Sponsor and the relevant agency/IRB/IEC with any additional requested information (e.g., autopsy reports and terminal medical reports). Any follow-up information on a previously reported SAE will also be reported within 24 hours of awareness.

If the Investigator does not have all information regarding an SAE, he/ she will not wait to receive additional information before reporting the event and completing the appropriate data collection form. The Investigator should always provide a preliminary assessment of causality at the time of the initial report as described in Section 10.5.3.

The primary mechanism for reporting SAEs will be a paper collection form.

Scan & email of the SAE form is the preferred method to transmit this information to the project contact for SAE receipt although fax transmission is also acceptable. In rare circumstances and in the absence of email facilities or fax equipment, notification by telephone is acceptable, with a copy of the SAE data collection form sent by overnight mail. Initial notification via the telephone does not replace the need for the Investigator to complete and sign the SAE data collection form within the designated reporting periods.

If required according to local regulatory authorities and IRB/IEC policy, each Investigator must then notify his or her relevant agency/IRB/IEC of the SAE.

NeRRe is required to expedite to worldwide regulatory authorities reports of SUSARs in line with the relevant legislation. Fatal or life-threatening SUSARs must be reported within 7 calendar days and other SUSARs within 15 calendar days.

In accordance with the European Commission Directive 2001/20/EC, NeRRe will notify the relevant Ethics Committees in concerned Member states of applicable SUSARs as individual notifications or through a periodic line listing. SUSARs will be reported to IRB/EC in non-EU countries according to local regulations and IRB/EC policy.

All investigators will receive a safety letter notifying them of relevant SUSAR reports.

NeRRe (or their designee) will submit to the regulatory authorities all safety updates and periodic reports, as required by applicable regulatory requirements.

11. STATISTICAL CONSIDERATIONS

11.1 Estimated Sample Size

11.1.1 Efficacy

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 hot flushes 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 will be determined by the adaptive dose-finding design algorithm (See Appendix B). In addition, the actual sample size is increased by 10% to allow for subjects who withdraw prematurely and so the initial sample size will be 165 subjects.

The sample size allows for 95% power since each of the four co-primary endpoints are required to demonstrate statistical significance in order to yield a firm conclusion of beneficial treatment effect on hot flushes. Conservatively, assuming independence among the four endpoints, this would yield approximately 81% power overall for achieving statistical significance for each of the four endpoints. Thus, overall power for the trial is estimated to be at least 81%; the actual power depends on the unknown magnitudes of correlations among the co-primary endpoints.

11.1.2 Safety

In addition to assessing efficacy, the study objectives also include further evaluation of the safety of a range of doses of NT-814 in women with menopausal symptoms. Regarding estimation of differences in AE rates from placebo for a given dose group, Table 11-1 (computed via normal approximation to the binomial distribution) shows the half-width of 95% confidence intervals for difference between a pair of treatment groups. This indicates the precision for such estimation for several example cases. The DRC will use these confidence intervals to assist in their consideration of increasing the sample size of those doses which achieve target levels of HF frequency and severity reduction in order to better assess safety. Criteria for such considerations will be described more fully in the DRC Charter.

Table 11-1 Estimates of 95% confidence interval half-widths for a representative range of treatment group differences in incidence of adverse events and sample sizes

OBSERVED between-group difference in percentage	95% confidence interval half-width for indicated N/group and OBSERVED percentage point difference from placebo		
	N=30	N=40	N=50
15 - 5 = 10	15	13	12
25 – 5 = 20	17	15	13
35 – 5 = 30	19	16	15
20 – 10 =10	18	15	14
30 – 10 = 20	20	17	15
40 – 20 = 20	21	18	16
30 – 20 = 10	22	19	17
40 – 20 = 20	23	20	18
50 – 20 = 30	23	20	18

11.2 Interim Reviews

A number of interim analyses (IA) are planned to enable the following two questions to be addressed as the trial progresses:

1) Do the emerging efficacy and safety data support the continued evaluation of all four NT-814 doses?

2) Do the assumptions used in the original sample size estimate remain valid?

The first IA will be conducted after efficacy data for the frequency and severity of hot flushes is collected through Week 4 from 30 patients randomized in equal proportions to the four NT-814 doses and placebo. The DRC will review the HF frequency and severity results, together with all available safety data, and will determine the randomisation ratios for subjects subsequent enrolled into the study. The ratio may be changed to optimise the adaptive design to estimate doses with target levels of weekly average daily HF reduction from baseline of between 6 and 8. The adaptive design algorithm will be based on Bayesian Emax dose-response modeling[57] and / or T-statistic adaptive dose-finding design[58]. The DRC will have dose assignment recommendations by each of those adaptive dose-finding algorithms and assign the randomization ratio that best integrates the safety and efficacy objectives of the trial and the adaptive design results. Additional details regarding the adaptive algorithms and their performance characteristics are in the adaptive dose-finding design simulation report (Appendix B). Subsequent IA's will be conducted similarly, after Week 4 data from each cohort of ~30 additional subjects become available. The DRC will evaluate stopping criteria according to the IA algorithms as described in Appendix B and decide on stopping assignment of subjects to specific doses or stopping the trial for an efficacy or safety conclusion, and dose choice for subsequent study or futility.

The DRC will be unblinded to treatment allocation. Subjects will continue to be recruited and randomised while the interim analyses are being undertaken.

11.3 Analysis Sets

The following data sets will be used for the statistical analysis.

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

2. **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 analyzed according to randomised treatment. This is the primary efficacy analysis set for the study.

3. **Per Protocol (PP) Set**: All subjects in the FAS excluding those identified as having relevant protocol deviations.

4. **Exposure-Response (ER) Set:** all subjects who received at least one dose of double-blind study drug and for whom the PK data are considered sufficient.

Analysis Sets will be identified prior to the unblinding of the study data.

11.4 Data Analysis

A Statistical Analysis Plan (SAP) will be written and finalised prior to database lock and treatment unblinding. The SAP will give a more detailed description of the summaries and analyses that will be performed, expanding on the protocol specified analysis. Any deviation from the protocol specified analysis will be documented within a protocol amendment or in the SAP, as appropriate, and described within the CSR.

Data will be analysed using SAS® version 9.4 or a later version.

Continuous variables will be summarised using number of non-missing observations, mean, standard deviation, median, first and third quartiles, minimum and maximum values. Categorical variables will be summarised using counts and percentages. All statistical hypothesis tests and confidence intervals will be two sided, using a type I error rate of 0.05. No adjustment for multiple comparisons will be used for this Phase 2 study. For the analysis of each efficacy endpoint pairwise statistical comparisons are planned for each NT-814 dose group (40mg, 80mg, 120 mg and 160 mg) versus placebo, with the understanding that this increases the overall type I error rate of the study. All collected subject data will be listed.

No subgroup analyses are planned to be performed.

11.5 Interim Analysis

The first interim analysis (IA) will be conducted after efficacy data from frequency and severity of hot flushes (HF) is collected through Week 4 from 30 patients randomized in equal proportions to the 4 doses and placebo. Based on DRC review of the IA results, randomization ratios for

subsequent enrolled patients may be changed to optimize the adaptive design to estimate doses with target levels of weekly average daily HF reduction from baseline 6 and 8. Adaptive design algorithm will be based on Bayesian Emax dose-response modeling[57] and / or T-statistic adaptive dose-finding design[58]. 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. Additional details regarding the adaptive algorithms and their performance characteristics are in the adaptive dose-finding design simulation report (Appendix B). 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, 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.

11.6 Protocol Deviations

Protocol deviations will be listed by subject and treatment group. Relevant protocol deviations will be identified from review of the subject data, prior to database lock, at a blinded data review meeting (BDRM). Subjects with any relevant deviations that are judged to impact on the statistical analysis will be excluded from the PP analysis set.

11.7 Subject Disposition

The number of subjects enrolled, randomised and within each subject analysis set will be summarised by treatment group and overall. Any screen failure subjects will be separately described in the CSR. In addition, the number of subjects completing / not completing the study will be summarised along with the primary reason for withdrawal from the study.

11.8 Demographics and Baseline Characteristics

Relevant Screening and Baseline data (i.e. data collected prior to the study treatment administration initiation) and demographic characteristics will be summarised descriptively for each treatment group. The FAS will be used in summaries of demographic and baseline data. No statistical testing will be used to compare treatment groups for different baseline characteristics.

11.9 Statistical Methods for Efficacy Parameters

All efficacy analyses will be performed on the FAS analysis sets which is the primary analysis set for all statistical comparisons of efficacy endpoints. The primary efficacy endpoint will, additionally, be analysed in the PP set.

11.9.1 Primary Efficacy Analysis

The co-primary endpoints are the change from baseline in the mean daily frequencies of moderate and severe hot flushes in the 7 days before the week 4 and Week 12 visits, and the mean severity of moderate and severe hot flushes in the 7 days before the week 4 and Week 12 visits. The baseline assessment will be the last 7 days of the baseline diary completion period. Absolute and

changes from baseline in the mean daily frequencies and average hot flush severity will be summarised by treatment group. The change from baseline endpoint will be analysed using a mixed model repeated measures analysis. Pairwise statistical comparisons are planned for each NT-814 dose group (40 mg, 80 mg, 120 mg and 160 mg) versus placebo, presenting the adjusted mean treatment difference with its corresponding 95% confidence interval. In addition, 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 [details will be in the Statistical Analysis Plan].

The primary objective will be assessed by testing the following hypotheses for each NT-814 group versus placebo separately:

Null hypothesis (H_0): There is no difference in the mean change from baseline in hot flush frequency [or severity] at Week 12 [or Week 4] compared to Baseline for the NT-814 treatment group compared to placebo.

H_0: μ (active) = μ (placebo)

Alternative hypothesis (H_1): There is a difference in the mean change from baseline in hot flush frequency [or severity] at Week 12 [or Week 4] compared to Baseline for the NT-814 treatment group compared to placebo.

H_1: μ (active) $\neq$ μ (placebo)

Absolute and changes from baseline in the in the mean frequency and severity will be summarised by treatment group.

11.9.2 Secondary Efficacy Analysis

11.9.2.1 Other hot flush frequency and severity secondary endpoints

Each of the hot flush frequency and severity secondary endpoints, including night time awakening (Section 8.2), will be summarised by treatment group and scheduled visit (as applicable) and will be analysed, following the same approach as for the primary endpoint.

The proportion of responders, defined as a reduction of >=50% points and >=80% points on the weekly average frequency of hot flushes will be summarised at each scheduled visit. The Week 4 and Week 12 response will be analysed by logistic regression, including the same terms in the model as for the primary efficacy endpoint analysis. Pairwise statistical comparisons are planned for each NT-814 dose group versus placebo, presenting the adjusted odds ratio for the treatment difference with its corresponding 95% Wald confidence interval.

11.9.2.2 Other Secondary endpoints

The analyses of the following endpoints will use the same approach as for the primary endpoint: Pittsburgh Sleep Quality Index scores, Insomnia Severity Index Scores, Beck Depression Inventory Scores, MenQoL and HFRDIS (total score and domain/sub-scores, as applicable).

For each of these secondary endpoints the appropriate analysis method to be used to assess the observed data will be reviewed, and any changes documented in the SAP. Only the Week 4 and Week 12 time points will be analysed, other visits summarised only, by treatment group.

11.10 Statistical Methods for Safety Parameters

The safety analysis set will be used for all presentations of safety endpoints. No statistical testing will be used to compare treatment groups for different safety endpoints. Safety data will be summarised descriptively for each treatment group.

11.10.1 Adverse Events

All adverse events will be coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA). Treatment-emergent adverse events (TEAE) will be presented within summary presentations, by MedDRA system organ class (SOC), preferred term and treatment group.

An adverse event will be defined as treatment emergent if the onset date is on or after the date of first dosing with study treatment. Any adverse event with an onset date earlier than the first dosing with study treatment will be considered as a pre-treatment adverse event. 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. Pre-treatment AEs will be identified in a subject listing.

Summaries presentations for the number and percentage of subjects reporting TEAEs, severity of TEAEs, TEAEs by relationship to study treatment, and treatment emergent serious adverse events (SAE) will be presented by treatment group and overall. In addition, SAEs and adverse events directly resulting in withdrawal from study will be listed.

11.10.2 Laboratory Parameters

Laboratory parameters (haematology, biochemistry, bone turnover markers and urinalysis) will be summarised over the scheduled visits by treatment group. Absolute values and changes from baseline will be summarised for the haematology, biochemistry and bone turn-over parameters. In addition, for haematology and biochemistry parameters, shift tables will be presented for changes in category (below, within or above normal range) from baseline to the subjects' worst post-baseline value ('worst' separately identified as either lowest or highest value observed on a subject for specific parameter being presented). Laboratory values outside the reference range will be identified in the subject listings.

A summary of abnormal clinically significant laboratory values will also be produced.

11.10.3 Vital Signs

Vital sign parameters (heart rate, systolic and diastolic blood pressure, temperature, weight, BMI, waist circumference) will be summarised (absolute values and changes from baseline) by treatment group and scheduled protocol visit.

11.10.4 ECG

ECGs will be assessed by the Investigator for abnormal findings. These will be categorised and summarised as normal, abnormal-not clinically significant or abnormal-clinically significant. A summary of abnormal clinically significant ECG findings will be produced.

ECG intervals will be recorded and summarised as absolute values and changes from baseline. Other ECG findings will be listed.

11.10.5 C-SSRS

Absolute values and changes from baseline will be summarised by treatment group and scheduled protocol visit.

11.10.6 Concomitant Medications

Concomitant therapies will be coded to their generic name and Drug Class using the current version of the WHO Drugs Dictionary. Medications will be assigned as being prior to study treatment or concomitant with study treatment, based on the start and stop dates of the medication and the study treatment. If the medication stop date is before the study treatment start date, the medication will be assigned as being prior to study treatment. In all other situations, the medication will be assigned as being concomitant with study treatment. Concomitant medications will be summarised by WHO Drug Class, generic name and treatment group.

11.11 Pharmacokinetics Analysis

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. 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.

11.12 Missing Data

Summary statistics and statistical analysis will be based primarily on non-missing values. Sensitivity analyses may be conducted to check the robustness of the analysis results under alternative assumptions with regards to missing data. Further details on the handling of missing values, including rules applied to incomplete questionnaires and any planned sensitivity analyses will be defined in the SAP.

12. END OF THE STUDY

The end of the study will be defined as the last subject's last visit.

13. ETHICS COMMITTEE REVIEW/INFORMED CONSENT

13.1 Institutional Review Board (IRB) / Independent Ethics Committee (IEC) and Relevant Authorities

The final study protocol, the subject information and consent form and any other subject facing materials (e.g. questionnaires, advertisements) will be approved by an appropriately constituted IRB/IEC. Approval will be received in writing before initiation of the study.

Clinical study authorisation will be obtained prior to initiation of the study from the relevant Regulatory Authority.

13.2 Ethical Conduct of the Study

The study will be performed in accordance with the local regulations, Good Clinical Practice (GCP) as described by the International Council for Harmonization (ICH), and the ethical principles that have their origins in the Declaration of Helsinki.

13.3 Informed Consent

For each study subject, informed consent will be obtained prior to any protocol-related activities being undertaken. As part of this procedure, the principal investigator or one of his/her associates will explain orally and in writing the nature, duration, purpose of the study, and the action of the study drug in such a manner that the subject is aware of the potential risks, inconveniences, or adverse effects that may occur. They will be informed that their medical records may be reviewed by appropriately qualified monitors of the Sponsor or a Sponsor Representative, and by auditors or regulatory authorities to ensure the accuracy of the details recorded as part of the study. They will be informed that they may withdraw from the study at any time without prejudice to further treatment. They will receive all information that is required by local regulations and ICH guidelines.

13.4 Protocol Amendments

Once approved by the applicable Regulatory Authorities and IRBs/IECs, the protocol must not be amended without approval by NeRRe. Unless an amendment is required to be implemented urgently in the interests of safety, or is deemed administrative by NeRRe, any amendments to the protocol must be authorised by the applicable Regulatory Authorities and IRB/IEC prior to implementation.

14. STUDY AND DATA MANAGEMENT

14.1 Monitoring

In accordance with applicable regulations, including GCP, and NeRRe or delegated CRO's Standard Operating Procedures (SOP), clinical research monitors will, prior to the start of the study, review with the site staff the protocol, study requirements, and the sites's responsibilities to satisfy regulatory, ethical, and NeRRe requirements. This review will also include a review of source documentation. A list of what is to be classed as 'source documentation' will be documented in the Study Monitoring Manual.

During the study the clinical monitor will periodically visit the site to verify:

- Data are authentic, accurate and complete
- Safety and right of the subjects are being protected
- IMP accountability
- AE and SAE reporting
- Study is conducted in accordance with the currently approved protocol and any other study agreements, GCP and all applicable regulatory requirements

14.2 Quality Assurance

To ensure compliance with GCP and all applicable regulatory requirements, the investigator[s]/institution[s] will permit study-related audits, IRB/IEC review, and regulatory inspection[s], providing direct access to source data/documents.

14.3 Data Recording

An eCRF will be used to capture subject data into a secure, validated database. Access to enter data in the eCRF will be limited to delegated and trained investigator site staff only.

Primary efficacy endpoint data and some secondary efficacy endpoint data will be recorded directly by subjects into the eDiary. In addition, eC-SSRS data will be recorded directly by subjects into an interactive voice response (IVR) system. Subject-entered diary and IVR data will not be subjected to source data verification.

14.3.1 Data to be Considered as Source Data

Source data may be defined as information from an original record or certified copy of the original record containing patient information for use in the trial. The information may include, but is not limited, to clinical findings, observations, or other activities necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents

(original records or certified copies) (ICH E2A Guideline). Examples of source data include subject identification and randomisation identification.

14.4 Confidentiality

The Investigator must assure that the subjects' anonymity is maintained. On all study documentation, with the exception of the consent form and subject ID logs, subjects will only be identified by their unique identification code and will not be referred to by name.

Study data will be handled with utmost discretion within the context of physician's confidentiality. Names of participating subjects and data generated as a result of this study will not be passed on to unauthorized persons. If original documents are to be sent outside of the research site they must be redacted to remove information which might potentially identify the subject.

NeRRe may transfer some data collected during the study to a different company or regulatory authority outside of the US or EU for the purpose of processing, review, analysis or storage. Whenever the subject's personal data is transferred, it will be kept confidential and secure, and will used only for the purpose for which it was collected.

Safety analysis samples collected during the study will be analysed at a central laboratory. Samples will be identified by the subjects' unique identification code only. All safety samples will be destroyed after the assays have been completed.

Blood samples for analysis of NT-814 concentrations will be shipped to Aptuit Srl, Italy for analysis. Samples will be identified by the subjects' unique identification code only. Following completion of the analysis, all samples will be destroyed.

14.5 Retention of Study Data

Following closure of the study, the Investigator must maintain all site study records, except for those required by local regulations to be maintained by someone else, in a safe and secure location. The records must be maintained to allow easy and timely retrieval, when needed (e.g., audit or inspection), and, whenever feasible, to allow any subsequent review of data in conjunction with assessment of the facility, supporting systems, and staff. Where permitted by local laws/ regulations or institutional policy, some or all of these records can be maintained in a format other than hard copy (e.g., microfiche, scanned, electronic). The Investigator must, however, ensure that all reproductions are legible and are a true and accurate copy of the originals, and meet accessibility and retrieval standards, including re-generating a hard copy, if required. Furthermore, the Investigator must ensure there is an acceptable back-up of these reproductions and that an acceptable quality control process exists for making these reproductions.

NeRRe will inform the Investigator of the time period for retaining these records to comply with all applicable regulatory requirements. The minimum retention time will be either 15 years or, if

a longer period if required by local institutional requirements or local laws or regulations, then the minimum will become that longer period.

The Investigator must notify NeRRe of any changes in the archival arrangements, including, but not limited to, archival at an off-site facility or transfer of ownership of the records in the event the Investigator leaves the site. The Investigator must not dispose of any records without prior approval from NeRRe.

14.6 Communication and Publication of Results

Where required by applicable regulatory requirements, an Investigator signatory will be identified for the approval of the CSR. The Investigator will be provided appropriate access to statistical tables, figures, and relevant reports and will have the opportunity to review the complete study results at NeRRe or another mutually agreeable location.

The original eCRFs and all data generated during the study under this protocol will become the property of the Sponsor.

Upon completion of the CSR, NeRRe will ensure public disclosure of the clinical trial research results according to the NeRRe's SOP and provide each Investigator with the full summary of the study results. The Investigator is encouraged to share the summary results with the study subjects, as appropriate.

It is the intent of all parties that the results of the study be published in a timely manner consistent with academic standards and with due consideration given to the protection of intellectual property rights. The Principal Investigator and NeRRe will be responsible for writing the proposed publication; this must happen with due diligence and with minimal delay.

Any proposed publication or presentation (including a manuscript, abstract or poster) originated by an Investigator for submission to a journal or scientific meeting should be sent to the Sponsor for review at least sixty days prior to submission. The Sponsor's comments on the proposed publication shall be considered in good faith by the authors. Sponsor may delay such submission by a maximum of six months if it reasonably believes that publication of results may compromise its intellectual property rights or else insist that such information or data is removed from the proposed publication. Publication of the results will not include confidential information without the permission of the Sponsor.

14.7 Indemnification

In the event of study-related damage or injuries, the clinical trial insurance of the Sponsor provides compensation for claims that arise in accordance with the regulatory requirements of the countries involved, except for claims that arise from willful misconduct or gross negligence. A copy of the country-specific insurance certificates will be held in the TMF and in the Investigator Site File.

15. REFERENCES

1. Joffe H, Massler A, Sharkey KM. Evaluation and management of sleep disturbance during the menopause transition. *Semin Reprod Med*. 2010;28(5):404-21.
2. Santoro N, Epperson CN, Mathews SB. Menopausal Symptoms and Their Management. *Endocrinol Metab Clin North Am.* 2015;44(3):497-515.
3. Williams RE, Levine KB, Kalilani L, et al. Menopause-specific questionnaire assessment in US population-based study shows negative impact on health-related quality of life. *Maturitas*. 2009 Feb 20;62(2):153-9.
4. Pachman DR, Jones JM, Loprinzi CL. Management of menopause-associated vasomotor symptoms: Current treatment options, challenges and future directions. *Int J Womens Health.* 2010;2:123-35.
5. Whiteley J, Wagner JS, Bushmakin A, et al. Impact of the severity of vasomotor symptoms on health status, resource use, and productivity. *Menopause*. 2013;20(5):518-24.
6. Williams RE, Kalilani L, DiBenedetti DB, et al. Healthcare seeking and treatment for menopausal symptoms in the United States. *Maturitas*. 2007;58(4):348-58.
7. Reed SD, Ludman EJ, Newton KM, et al. Depressive symptoms and menopausal burden in the midlife. *Maturitas*. 2009;62(3):306-10.
8. Woods NF, Hohensee C, Carpenter JS, et al. Symptom clusters among MsFLASH clinical trial participants. Menopause. 2016; 23(2): 158-65
9. Stahl SM. Vasomotor symptoms and depression in women, part I. Role of vasomotor symptoms in signaling the onset or relapse of a major depressive episode. *J Clin Psychiatry*. 2009;70(1):11-2.
10. Worsley R, Bell RJ, Gartoulla P, et al. Moderate-severe vasomotor symptoms are associated with moderate-severe depressive symptoms. J of Womeens health 2017. DOI:1089/jwh.2016.6142
11. Ratner S, Ofri D. Menopause and hormone replacement. Part 1. Evaluation and treatment. *West J Med.* 2001; 174(6): 400–404.
12. Thurston RC, and Joffe H. Vasomotor Symptoms and Menopause: Findings from the Study of Women's Health Across the Nation. *Obstet Gynecol Clin North Am.* 2011 Sep; 38(3): 489–501.
13. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause: evidence from the Penn Ovarian Aging Study cohort. *Menopause*. 2014;21(9):924-32.
14. Avis NE, Crawford SL, Greendale G, et al. Study of Women's Health Across the Nation. Duration of menopausal vasomotor symptoms over the menopause transition. *JAMA Intern Med.* 2015;175(4):531-9.
15. MHRA/CHM (Drug Safety Update 2007; 1 (2): 2–6) available at www.gov.uk/drug-safety-update
16. Pinkerton JV, Pan K, Abraham L, et al. Sleep parameters and health-related quality of life with bazedoxifene/conjugated estrogens: a randomized trial. Menopause 2014 Mar;21(3):252-9
17. Santoro N, Allshouse A, Neal-Perry G, et al. Longitudinal changes in menopausal symptoms comparing women randomized to low-dose oral conjugated estrogens or transdermal estradiol plus micronized progesterone versus placebo: the Kronos Early Estrogen Prevention Study. Menopause. 2017 Mar;24(3):238-246.
18. Rance NE, Dacks PA, Mittelman-Smith MA, et al. Modulation of body temperature and LH secretion by hypothalamic KNDy (kisspeptin, neurokinin B and dynorphin) neurons: a novel hypothesis on the mechanism of hot flushes. *Front Neuroendocrinol*. 2013;34(3):211-27.
19. Casper RF, Yen SSC & Wilkes MM. Menopausal flushes. A neuroendocrine link with pulsatile luteinising hormone secretion. Science 1979; 205: 823-825.
20. Casper RF, Yen SS. Menopausal flushes: effect of pituitary gonadotropin desensitization by a potent luteinizing hormone-releasing factor agonist. J Clin Endocrinol Metab. 1981;53(5):1056-8
21. Casper RF, Yen SS. Neuroendocrinology of menopausal flushes: an hypothesis of flush mechanism. Clin Endocrinol (Oxf). 1985; 22: 293-312.
22. DeFazio J, Meldrum DR, Laufer L, et al. Induction of hot flashes in premenopausal women treated with a long-acting GnRH agonist. J Clin Endocrinol Metab. 1983;56:445-8
23. Gambone J, Meldrum DR, Laufer L, et al. Further delineation of hypothalamic dysfunction responsible for menopausal hot flashes. J Clin Endocrinol Metab. 1984;59(6):1097-102.

24. Lehman MN, Coolen LM, Goodman RL. Minireview: Kisspeptin/Neurokinin B/Dynorphin (KNDy) Cells of the Arcuate Nucleus: A Central Node in the Control of Gonadotropin-Releasing Hormone Secretion. Endocrinology. 2010;151(8):3479–3489.
25. Rance NE, McMullen NT, Smialek JE, et al. Postmenopausal hypertrophy of neurons expressing the estrogen receptor gene in the human hypothalamus. J Clin Endocrinol Metab. 1990;71:79-85
26. Rance NE, Young WS., III Hypertrophy and increased gene expression of neurons containing neurokinin-B and substance-P messenger ribonucleic acids in the hypothalami of postmenopausal women. Endocrinology. 1991;128(5):2239–2247
27. Rometo AM, Rance NE. Changes in prodynorphin gene expression and neuronal morphology in the hypothalamus of postmenopausal women. J Neuroendocrinol. 2008;20(12):1376-81
28. Abel TW, Voytko ML, Rance NE. The effects of hormone replacement therapy on hypothalamic neuropeptide gene expression in a primate model of menopause. J Clin Endocrinol Metab. 1999;84(6):2111-8
29. Sandoval-Guzmán T, Stalcup ST, Krajewski SJ, et al. Effects of ovariectomy on the neuroendocrine axes regulating reproduction and energy balance in young cynomolgus macaques. J Neuroendocrinol. 2004;16(2):146-53
30. Rometo AM, Krajewski SJ, Voytko ML, Rance NE. Hypertrophy and increased kisspeptin gene expression in the hypothalamic infundibular nucleus of postmenopausal women and ovariectomized monkeys. J Clin Endocrinol Metab. 2007;92(7):2744-50.Boulant 2000
31. Rance NE, Dacks PA, Mittelman-Smith MA, et al. Modulation of body temperature and LH secretion by hypothalamic KNDy (kisspeptin, neurokinin B and dynorphin) neurons: a novel hypothesis on the mechanism of hot flushes. Front Neuroendocrinol. 2013;34(3):211-27
32. Mittelman-Smith MA, Williams H, Krajewski-Hall SJ, et al. Role for kisspeptin/neurokinin B/dynorphin (KNDy) neurons in cutaneous vasodilatation and the estrogen modulation of body temperature. Proc Natl Acad Sci U S A. 2012;109(48):19846-51.
33. Jayasena CN, Comninos AN, Stefanopoulou E, et al. Neurokinin B administration induces hot flashes in women. Sci Rep. 2015;5:8466
34. Prague JK, Roberts RE, Comninos AN, et al. Neurokinin 3 receptor antagonism as a novel treatment for menopausal hot flushes: a phase 2, randomised, double-blind, placebo-controlled trial. Lancet 2017;389:1809-1820.
35. Fraser GL, Depypere H, Timmerman D, et al. Clinical Evaluation of the NK3 Receptor Antagonist Fezolinetant (a.k.a. ESN364) for the Treatment of Menopausal Hot Flashes. Abstract OR16-5, Presented at ENDO 2017
36. Navarro VM, Bosch MA, León S, et al. The Integrated Hypothalamic Tachykinin-Kisspeptin System as a Central Coordinator for Reproduction. Endocrinology. 2015;156::627-637. doi:10.1210/en.2014-1651.
37. Arisawa M, De Palatis L, Ho R, et al. Stimulatory role of substance P on gonadotropin release in ovariectomized rats. Neuroendocrinology. 1990;51(5):523–529.
38. Yamamura T, Wakabayashi Y, Ohkura S, et al. Effects of intravenous administration of neurokinin receptor subtype-selective agonists on gonadotropin-releasing hormone pulse generator activity and luteinizing hormone secretion in goats. The Journal of Reproduction and Development. 2015;61(1):20-29. doi:10.1262/jrd.2014-109.
39. Kalil B, Ramaswamy S, Plant TM. The distribution of substance P and kisspeptin in the mediobasal hypothalamus of the male rhesus monkey and a comparison of intravenous administration of these peptides to release GnRH as reflected by LH secretion. Neuroendocrinology. 2016;103(6):711-723. doi:10.1159/000442420.
40. Hrabovszky E, Borsay BÁ, Rácz K, et al. Substance P Immunoreactivity Exhibits Frequent Colocalization with Kisspeptin and Neurokinin B in the Human Infundibular Region. Wolfe A, ed. PLoS ONE. 2013;8(8):e72369.
41. Borsay B, Skrapits K, Herczeg L, et al. Hypophysiotropic gonadotropin-releasing hormone projections are exposed to dense plexuses of kisspeptin, neurokinin B and substance p immunoreactive fibers in the human: a study on tissues from postmenopausal women. Neuroendocrinology 2014;100(2-3):141-52

42. Wong BJ, Minson CT. Neurokinin-1 receptor desensitization attenuates cutaneous active vasodilatation in humans. J Physiol 2006; 577:1043-51
43. Schaffalitzky De Muckadell OB, Aggestrup S, Stentoft P. Flushing and plasma substance P concentration during infusion of synthetic substance P in normal man. Scand J Gastroenterol. 1986 May;21(4):498-502.
44. Kramer MS, Cutler N, Feighner J, et al. Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science. 1998;281(5383):1640-1645.
45. Kramer MS, Winokur A, Kelsey J, et al. Demonstration of the efficacy and safety of a novel substance P (NK1) receptor antagonist in major depression. Neuropsychopharmacology. 2004;29(2):385-392.
46. Ratti E, Bellew K, Bettica P, et al. Results from 2 randomized, double-blind, placebo-controlled studies of the novel NK1 receptor antagonist casopitant in patients with major depressive disorder. J Clin Psychopharmacol. 2011;31(6):727-733
47. Ratti E, Bettica P, Alexander R, et al. Full central neurokinin-1 receptor blockade is required for efficacy in depression: evidence from orvepitant clinical studies. J Psychopharmacol. 2013;27(5):424-434.
48. Newton KM, Carpenter JS, Guthrie KA, et al. Methods for he design of vasomotor symptoms trials; the MsFLASH Network. Menopause 2014 21(1) doi:10.1097/GME.0b013e31829337a4.
49. FDA guidance for industry: oestrogen and oestrogen/progestin drug products to treat vasomotor symptoms and vulvar and vaginal atrophy symptoms – recommendations for clinical evaluation. January 2003.
50. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989 May;28(2):193-213.
51. Hilditch JR, Lewis J, Peter A, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. Maturitas. 1996; 24:161–175.
52. Carpenter JS. The Hot Flash Related Daily Interference Scale: a tool for assessing the impact of hot flashes on quality of life following breast cancer. J Pain Symptom Manage. 2001; 22:979–989.
53. Carpenter JS, Storniolo AM, Johns S, et al. Randomized, double-blind, placebo-controlled crossover trials of venlafaxine for hot flashes after breast cancer. Oncologist. 2007; 12:124–135.
54. Beck AT, Steer RA and Brown GK (1996) "Manual for the Beck Depression Inventory-II". San Antonio, TX: Psychological Corporation
55. CHMP Guideline on clinical investigation of medicinal products for hormone replacement therapy of oestrogen deficiency symptoms in postmenopausal women. May 2006.
56. Simon JA, Portman DJ, Kaunitz AM, et al. Low-dose paroxetine 7.5 mg for menopausal vasomotor symptoms: two randomized controlled trials. Menopause 2013;20:1027-1035.
57. COMPASS 1.1 (E) User Manual (2012), Cytel Inc., Cambridge, MA, USA.
58. 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.

16. SIGNATURES AND AGREEMENT WITH THE PROTOCOL

Sponsor Approval

I have reviewed and approved the protocol and confirm that the protocol follows GCP.

Signature: _____ Date: 07 Feb 2019

Stephen Pawsey MBBS FRCA FFPM
Chief Medical Officer, NeRRe Therapeutics Ltd.

Principal Investigator Agreement

I agree to conduct the study according to the terms and conditions of this protocol, current Good Clinical Practice and with applicable regulatory requirements. All information pertaining to the study shall be treated in a confidential manner.

Signature: _____

Name of Principal Investigator:

Title:

Date:

17. APPENDICES

APPENDIX A: LIST OF PROHIBITED CONCOMITANT MEDICATIONS

The following is a comprehensive (but not exhaustive list*) of prohibited concomitant medications.

** Recently approved drugs may not be included in the list and should be checked on a case-by-case basis.*

Prohibited from the time shown until 1 week after the last dose of study medication	
Medications potentially confounding efficacy	**Hormonal therapies** • Vaginal hormonal products (rings, creams, gels and including DHEA or analogues thereof) – from 1 week prior to Screening Visit 2 • Transdermal oestrogen alone or or oestrogen/progestin products - from 4 weeks prior to Screening Visit 2 • Oral oestrogen (including selective oestrogen receptor modulators) and/or progestin therapy - – from 8 weeks prior to Screening Visit 2 • Intrauterine progestin therapy – from 8 weeks prior to Screening Visit 2 • Progestin implants and oestrogen alone injectable drug therapy – from 3 months prior to Screening Visit 2 • Oestrogen pellet therapy or progestin injectable drug therapy - – from 6 months prior to Screening Visit 2 **Non-hormonal therapies** Non-hormonal prescription (eg paroxetine, other anti-depressants including selective serotonin reuptake inhibitors, serotonin norepinephrine reuptake inhibitors and tri-cyclic antidepressants, alpha agonists [clonidine], methyldopa, gabapentin, pregabalin, medicinal cannabis or derivatives) or over the counter/herbal treatments for treatment of menopausal symptoms is not allowed throughout the study. Subjects must have discontinued these drugs at least 28 days prior to Screening Visit 2. Subjects may, however, be permitted to continue to use these drugs if the dose has been stable for at least 4 weeks and they have been prescribed solely for the management of another disorder (e.g. neuropathic pain, depression).
Prohibited from Screening Visit 2 until 1 week after the last dose of study medication	
CYP3A4 substrates with a narrow therapeutic range	alfentanil, astemizole, cisapride, cyclosporine, dihydroergotamine, ergotamine, fentanyl, pimozide, quinidine, sirolimus, tacrolimus, terfenadine
Strong or moderate	clarithromycin, atazanavir, idelalisib, nefazodone, nelfinavir,

inhibitors of CYP3A4	boceprevir, cobicistat, danoprevir and ritonavir, elvitegravir and ritonavir, grapefruit juice, indinavir and ritonavir, itraconazole, ketoconazole, lopinavir and ritonavir, paritaprevir and ritonavir and (ombitasvir and/or dasabuvir), posaconazole, ritonavir, saquinavir and ritonavir, telaprevir, tipranavir and ritonavir, troleandomycin, voriconazole aprepitant, cimetidine, ciprofloxacin, clotrimazole, crizotinib, cyclosporine, dronedarone, fluconazole, fluvoxamine, imatinib, tofisopam, verapamil, troleandomycin, ciprofloxacin
P-glycoprotein inhibitors	azithromycin, conivaptan, cyclosporine, diltiazem, erythromycin, felodipine, ketoconazole, quinidine, ranolazine, amiodarone, carvedilol, clarithromycin, dronedarone, itraconazole, lapatinib, propafenone, quinidine,
P-glycoprotein substrates with a narrow therapeutic range	Digoxin
Prohibited from Screening Visit 2 until Week 12	
Strong or moderate inducers of CYP3A4	rifampicin, carbamazepine, efavirenz, bosentan, modafinil, St. John's Wort, enzalutamide, mitotane, phenytoin, etravirine

APPENDIX B: PHASE 2B ADAPTIVE DOSE-FINDING DESIGN FOR NT-814 EFFECT ON REDUCTION OF HOT FLUSHES

J.Bolognese
18 July 2018

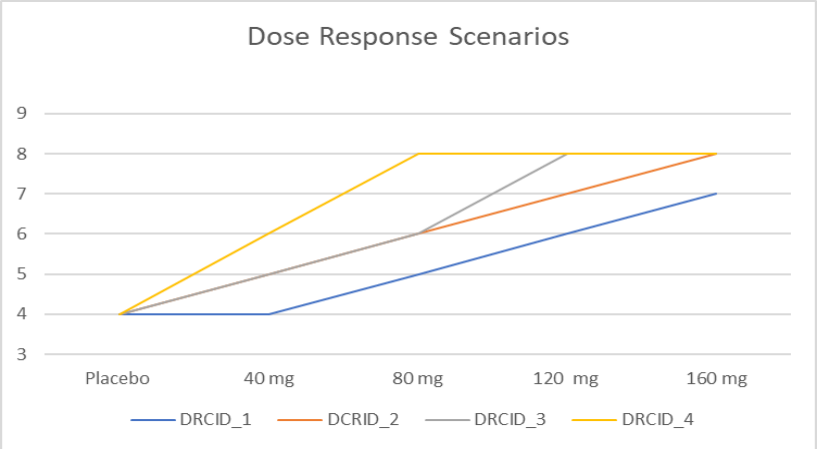
INTRODUCTION AND PURPOSE

A Phase 2b dose-finding placebo-controlled trial of NT-814 treatment with four different doses for 12 weeks aimed at reduction of hot flushes is being planned. Adaptive design options are being considered to modify randomization ratios to the NT-814 doses based on accumulating data. This report summarizes several example adaptive design options along with their performance characteristics for an initial assumed example set of potential TRUE underlying dose-response curves. The primary endpoints for the trial are hot flush frequency and severity; however, the adaptive design simulation software cannot handle multiple endpoints. Hence, the scenarios evaluated focus on reduction of hot flush frequency.

METHODS

First, in order to arrive at some example TRUE underlying dose-response scenarios, prior data on hot flush reduction was examined. NT-814 has shown efficacy in reduction of hot flushes over a 2-week period in a Phase 1/2a trial. Mean reductions in hot flush frequency of up to 9 per day from a baseline frequency ~10-12 were observed (with an observed SD of change from baseline ranged between ~3.4 to ~4.9; SD back-calculated from SE's from the analysis model yielded SD's of change ~3.7). In order to obtain longer term prior data to inform on expected variability, the Duavee label was also examined. It yielded SD's of change ~4.4 after 4 weeks' treatment and ~3.6 after 12 weeks' treatment. Based on these prior data, an SD=4.4 and somewhat conservative magnitudes of treatment effect were used to simulate adaptive design outcomes and performance characteristics. The following TRUE underlying dose-response curves were examined:

	Assumed TRUE underlying mean hot flush frequency reduction from baseline (SD=4)				
label	placebo (dose0)	dose1	dose2	dose3	dose4
DRCID_1	4	4	5	6	7
DRCID_2	4	5	6	7	8
DRCID_3	4	5	6	8	8
DRCID_4	4	6	8	8	8



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 hot flushes 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 will be determined by the adaptive dose-finding design algorithm.

The sample size allows for 95% power since each of the four co-primary endpoints are required to demonstrate statistical significance in order to yield a firm conclusion of beneficial treatment effect on hot flushes. Conservatively, assuming independence among the four endpoints, this would yield approximately 81% power overall for achieving statistical significance for each of the four endpoints. Thus, overall power for the trial is estimated to be at least 81%; the actual power depends on the unknown magnitudes of correlations among the co-primary endpoints.

Adaptive dose-finding design algorithms use results from interim analysis (IA) of accumulating data from the trial to modify dose-choices and / or randomization ratios for subsequent subjects entering the trial. Three numbers of IA and two adaptive dose-finding algorithms were evaluated for each of the four potential TRUE underlying dose-response curves (total 24 scenarios = 4 DR curves X 3 cohort sizes X 2 design algorithms). The three IA sizes evaluated are (i) no IA's (i.e., N=15 per dose group to compare to a non-adaptive design), (ii) two IAs, and (iii) five IAs. Both the 2-IA and 5-IA designs randomize the initial 25 patients 1:1:1:1:1 to the 5 dose groups (note, placebo is the zero dose group) to obtain a reasonable amount of seeding data upon which to begin the adaptation. Note, the actual planned size of the first cohort is N=30, which would yield results between those of the two sample size scenarios simulated for the adaptive designs. Hence, the actual performance characteristics of the trial fall between the those two sample size scenarios. Since the timing of the actual IA's may vary due to logistical / enrolment constraints, these two scenarios are representative of the those expected during the trial. Then the 2-IA design carries out a second IA after N=50 total patients are observed, and then the final analysis is

carried out after the 75 patients are completed. After the initial IA at N=25, the 5-IA design carries out an IA after results of each 10 patients are added to the accumulating data.

The two adaptive algorithms are (i) a Bayesian adaptive Emax (S-shaped) dose-response model fit at each IA (COMPASS User Manual), and (ii) a model based on classical T-statistics at each IA (Ivanova, 2008). The Emax design uses a "vague" Bayesian prior distribution on each of the four parameters of the S-shaped Emax model in combination with the observed data to estimate the posterior distribution of the Emax model parameters from which estimates of mean response at each dose are obtained, along with their SE's. Then randomization ratios are chosen for the next cohort to minimize the variances of the resulting Emax model estimates at two target levels of response, namely hot flush reductions of 6 per day and of 8 per day. The Emax algorithm was set up to recommend stopping randomization to a given dose if (a) the probability is > 90% that the mean hot flush reduction 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).

The T-statistic design relies on isotonic regression (i.e., best-fit non-decreasing dose-response curve) derived estimated means for each dose and then computes the resultant T-statistic for estimated mean difference from each target level of response (hot flush reduction of 6 and of 8) for each dose. Then the subjects in the next cohort are split among the two doses with T-statistics pointing closest to each target level of response and placebo. The sample sizes of each cohort are always 10 on placebo for the 2 IA design, and always 5 on placebo for the 5 IA design, to yield N=30 on placebo for the final analysis. For the simulations, the subjects assigned to the active doses are randomized according to the randomization ratio indicated by the adaptive design algorithm for each cohort after the first. The only stopping criteria recommendation available for the T-statistic in the simulation software (Cytel's COMPASS V1.1) is futility stopping; hence, if probability of a significant difference from placebo is <5% for a given dose based on the IA results, then the algorithm recommends that dose is dropped from any further randomization to it.

The 24 scenarios (4 DR curves X 3 cohort sizes X 2 design algorithms) were each simulated 500 times, and the following performance characteristics, averaged across the 500 simulations, were summarized in Tables 1a and 1b

- Average total sample size
- Power to reject a flat dose-response curve (alpha=0.05 2-sided), and conclude statistically significant NT-814 effect
- Average sample size assigned to each dose
- Average estimated dose level (among doses 1,2,3,4) closest to the target hot flush reduction from baseline of 6
- Percent of simulations choosing the correct dose with TRUE underlying hot flush reduction from baseline of 6
- Percent of simulations choosing the correct dose or adjacent to the correct dose with TRUE underlying hot flush reduction from baseline of 6
- Average estimated dose level (among doses 1,2,3,4) closest to the target hot flush reduction from baseline of 8
- Percent of simulations choosing the correct dose with TRUE underlying hot flush reduction from baseline of 8
- Percent of simulations choosing the correct dose or adjacent to the correct dose with TRUE underlying hot flush reduction from baseline of 8

In order to account for recruitment rate and lag of data collection and analysis as patients continually enter the trial, the simulations were run assuming a 6-week lag for each patient entering and their 4 week hot flash frequency data entering the interim analysis. An average recruitment rate of 5 patients per week was assumed, since the software cannot accommodate varying recruitment.

It should be noted that the analysis models for the actual trial data will be longitudinal mixed effects models to model the correlation structure of all hot flash frequency data at each week from baseline to week 12. Due to software limitations, the simulations only assume available data of Week 4, so the performance characteristics in this report are likely conservative.

SIMULATION RESULTS

For dose-response curve 1 (DRCID_1), power is approximately 90% for both the adaptive and the non-adaptive designs - highest for the (Emax 90 & 93%; Tstat 85 & 89%; non-adaptive design (88 & 89%). The adaptive designs achieve this power with slightly smaller overall sample sizes (~148 for Emax, ~145 for Tstat), since they have potential for early stopping for futility. Note that all of the adaptive designs, except the Emax 3-cohort design, assign fewer subjects to the lower 2 doses that have TRUE underlying mean hot flash values less than the lower target value 6, and assign the majority of subjects to doses 3 and 4, the former having TRUE response equal to the lower target value 6. The Emax 3-cohort design assigns more subjects to Dose1 (which has TRUE underlying response 4) than to doses 2 and 3 in order to best fit the lower portion of the Emax model. All of the adaptive designs assign most subjects to Dose 4. The subtle distinction between the Emax and Tstat adaptive designs (more spread of subjects across the dose range via Emax design than via Tstat design) arises since the Emax design estimates target doses via optimally fitting the Emax model, whereas the Tstat design seeks to assign most subjects to the two doses closest to the target levels of response. The average estimated target doses for both designs are very close to the TRUE target values (>4 for the upper target and approximately dos e3 for the lower target. However, only approximately 2/3 of the simulations actually estimate the target dose exactly at Dose3 for the lower target, but nearly all simulations estimate the lower target dose at Dose3, or Dose4 (i.e., at or adjacent to the correct / TRUE target dose).

For dose response curves 2,3 and 4 (DRCID_2, DRCID_3, DRCID_4) power is at least 94% for the adaptive designs, and generally similar to that for the non-adaptive 1-cohort designs. The Emax design achieves this power with somewhat smaller average sample sizes (N=133-146) than the Tstat and non-adaptive designs. The Tstat design sample sizes are N=146-149 because in a small portion of trials the design incorrectly stops early for futility - note the software does not permit inclusion of early stopping for an efficacy conclusion [NOTE: in the actual trial, the same early stopping criteria can be employed with the Tstat design as with the Emax design; it is just that the simulation software does not have that feature for the Tstat design]. Note also that the Emax design assigns more subjects to dose 4 (the highest dose) in order to fit the plateau response level, whereas the Tstat design focuses the dose assignments more at the doses with TRUE target levels of response 6 and 8. The Tstat design seems slightly better than the Emax model in estimating the target doses on average; Tstat design average estimated target doses are generally closer to TRUE target doses or mid-point of TRUE target doses when multiple doses have same TRUE target level of response. The percentage of correct target dose estimates are generally higher for the Tstat design than for the Emax design, except for the DRCID_2 lower target dose, for which they are generally similar, but the percentages are generally similar for the Emax adaptive designs compared to the non-adaptive designs, but higher for the Tstat design than

for the Emax and non-adaptive designs. For the Emax design, this is could be at least partially due to the early stoppages of the Emax design indicated by lower average sample sizes.

Type 1 error (false positive, i.e., rejecting the null hypothesis of flat dose-response curve when the TRUE underlying dose-response curve is flat) was estimated via simulation in the same manner as performance characteristics were simulated and summarized for the above four non-flat dose-response curves (performance characteristics are summarized in Table 2). Type 1 error is estimated to be approximately 5% for both adaptive designs and the non-adaptive designs analyzed by the same methods as the adaptive designs. Hence, the adaptive designs adequately preserve type 1 error. Note the much lower average sample sizes associated with the adaptive designs; they result from the high proportion of adaptive designs that stop early for a futility conclusion (66-78%). Finally, note the adaptive designs assign the largest number of subjects to dose 4, and minimize assignment to the lower doses.

RECOMMENDATIONS for consideration

Simulation software limitations to a single endpoint at a single time point make the performance characteristics summarized conservative. The actual analysis model can incorporate all time points of observation and multiple endpoints (frequency and severity); hence, actual statistical precision and performance characteristics will likely improve during the actual study beyond those levels reported in this simulation report. An adaptive dose-finding design is preferable to the fixed randomization design since it will allocate more subjects to doses with target levels of response and/or minimize randomization to doses below or above those with target levels of response. In addition, both the Emax and Tstat algorithms can be run at each IA and the suggested dose-assignments can be compared by the Data Review Committee (DRC), which can recommend the best dose ratio assignments in the context of the overall accumulated data available at the IA. Alternatively, depending on the performance characteristic(s) of most interest, the better performing adaptive design on that(those) performance characteristic(s) can be chosen.

REFERENCES

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

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

TABLE 1a - Performance Characteristics (sample sizes, power, early stopping probability)

TRUE underlying Dose-Response Curve	Design	Number of Cohorts	Average Total Sample Size	Power (%)*	Average Sample Size Assigned Per Dose					Proportion stopped early
					D0 (pbo)	D1	D2	D3	D4	
D0=4, D1=4, D2=5, D3=6, D4=7	Emax	1	150	88	30.0	30.0	30.0	30.0	30.0	0.000
		3	147	90	29.5	25.1	18.4	20.5	53.9	0.028
		6	148	93	29.6	21.6	19.1	22.2	55.3	0.026
	Tstat	1	150	89	30.0	30.0	30.0	30.0	30.0	0.000
		3	144	85	28.8	13.9	23.2	33.3	44.8	0.074
		6	145	89	29.0	15.2	24.4	32.4	44.2	0.066
D0=4, D1=5, D2=6, D3=7, D4=8	Emax	1	150	98	30.0	30.0	30.0	30.0	30.0	0.000
		3	146	99	29.1	30.2	22.3	20.8	43.1	0.056
		6	145	99	29.0	28.1	23.7	21.9	42.4	0.078
	Tstat	1	150	97	30.0	30.0	30.0	30.0	30.0	0.000
		3	146	94	29.3	21.1	30.8	34.0	31.2	0.038
		6	148	96	29.6	24.3	31.7	32.3	30.1	0.026
D0=4, D1=5, D2=6, D3=8, D4=8	Emax	1	150	99	30.0	30.0	30.0	30.0	30.0	0.000
		3	142	98	28.4	31.5	22.9	20.6	38.6	0.112
		6	141	100	28.2	29.7	24.2	21.5	37.2	0.150
	Tstat	1	150	99	30.0	30.0	30.0	30.0	30.0	0.000
		3	148	97	29.6	23.0	35.1	33.4	26.9	0.020
		6	149	98	29.7	25.8	36.4	30.8	25.9	0.016
D0=4, D1=6, D2=8, D3=8, D4=8	Emax	1	150	96	30.0	30.0	30.0	30.0	30.0	0.000
		3	136	95	27.1	39.2	20.1	16.6	32.5	0.186
		6	133	96	26.7	38.1	20.3	17.6	30.6	0.242
	Tstat	1	150	97	30.0	30.0	30.0	30.0	30.0	0.000
		3	146	94	29.2	38.7	34.4	23.6	20.1	0.044
		6	149	97	29.8	42.7	34.4	23.4	18.7	0.010

***Power to yield statistically significantly increasing trend across doses (i.e., not flat) at alpha=0.025 1-sided**

Table 1b - Performance Characteristics (accuracy of target dose estimation)

TRUE underlying Dose-Response Curve	Design	# of Cohorts	Dose with Target Response 8				Dose with Target Response 6			
			TRUE Target Dose	Average Estima-ted Target Dose	% correct Target Dose Estimates	% correct or nearly correct Target Dose Estimates	TRUE Target Dose	Average Estim-ated Target Dose	% correct Target Dose Estimates	% correct or nearly correct Target Dose Estimates
D0=4, D1=4, D2=5, D3=6, D4=7	**Emax**	**1**	>4	4.7	86	100	3	3.2	58	97
		3	>4	4.7	89	99	3	3.1	59	97
		6	>4	4.6	89	99	3	3.0	62	96
	Tstat	**1**	>4	4.5	73	100	3	3.0	71	99
		3	>4	4.5	74	98	3	3.0	66	97
		6	>4	4.6	76	99	3	3.0	68	98
D0=4, D1=5, D2=6, D3=7, D4=8	**Emax**	**1**	4	4.2	54	100	2	1.9	66	99
		3	4	4.1	55	96	2	1.9	67	99
		6	4	4.0	55	98	2	1.9	71	99
	Tstat	**1**	4	4.0	57	100	2	2.0	68	100
		3	4	3.9	54	100	2	2.0	65	99
		6	4	3.9	57	100	2	2.0	64	99
D0=4, D1=5, D2=6, D3=8, D4=8	**Emax**	**1**	3(,4)	3.8	33	81	2	1.7	66	100
		3	3(,4)	3.8	26	78	2	1.7	65	98
		6	3(,4)	3.8	26	81	2	1.7	67	98
	Tstat	**1**	3(,4)	3.6	47	86	2	1.8	76	100
		3	3(,4)	3.5	54	88	2	1.9	76	99
		6	3(,4)	3.5	50	88	2	1.9	78	99

D0=4, D1=6, D2=8, D3=8, D4=8	**Emax**	**1**	2(,3,4)	3.4	22	51	1	0.8	93	100
		3	2(,3,4)	3.3	18	52	1	0.9	85	100
		6	2(,3,4)	3.2	21	54	1	0.9	88	100
	Tstat	**1**	2(,3,4)	3.0	39	63	1	1.0	91	100
		3	2(,3,4)	2.9	42	66	1	1.0	88	100
		6	2(,3,4)	2.9	43	68	1	1.0	93	100

TABLE 2 - Performance Characteristics for TRUE underlying FLAT dose-response curve (sample sizes, power, early stopping probability); i.e., Hot Flash reduction 4 for all doses and placebo.

TRUE underlying Dose-Response Curve	Design	Number of Cohorts	Average Total Sample Size	Power (%)*	Average Sample Size Assigned Per Dose					Proportion stopped early
					D0 (pbo)	D1	D2	D3	D4	
D0=4, D1=4, D2=4, D3=4, D4=4	Emax	1	150	5.00	30.0	30.0	30.0	30.0	30.0	0.000
		3	101	4.60	20.2	22.8	11.1	11.5	35.6	0.674
		6	97	5.40	19.3	20.3	12.3	12.7	32.1	0.784
	Tstat	1	150	4.20	30.0	30.0	30.0	30.0	30.0	0.000
		3	103	4.60	20.6	10.3	13.2	18.8	40.0	0.664
		6	107	5.00	21.5	12.3	16.7	18.9	38.0	0.682

APPENDIX C: ASSESSMENT PROFORMAS

Note: The versions of the assessments provided in this Appendix are for reference only and may differ slightly from the actual versions used on the study. Where appropriate, permission and licenses to use the assessments will be obtained.

Hot Flush Screening Diary (Paper Diary)

Site Number _____ SWITCH-1 STUDY Subject Number _____

ONCE DAILY HOT FLUSH SCREENING DIARY

Please fill out this form **once a day** for 7 days.

Each evening, at bedtime, please record the total number of hot flushes of each severity you had in the last 24 hours.

Complete this section in the evening (before bedtime)	
Date:	__ __/ __ __ __/ __ __ D D M M M Y Y
No hot flushes during the last 24 hours	O
Total **number** of hot flushes of **each severity** during the last 24 hours: **MILD**: *Sensation of heat without sweating* **MODERATE**: *Sensation of heat with sweating, but able to continue activity.* **SEVERE**: *Sensation of heat with sweating, causing cessation (stopping) of activity.* **Note**: *A night-time hot flush that wakes you up (stops you sleeping) is a severe hot flush.* *If you had more hot flushes once you were awake, record these too.*	⌊⌋⌋ Mild ⌊⌋⌋ Moderate ⌊⌋⌋ Severe

Hot Flush Electronic Diary

The design of the electronic diary will be based around the following instructions and questions:

Instructions

Please complete this eDiary **twice a day** each day you are in the study (starting the evening of Screening Visit 2).

Each evening, at bedtime, please record the total number of hot flushes of each severity you had during that day since waking. Also, please confirm whether you took your study medication.

Each following morning, upon waking, please record the number of times you woke up in the night, and the total number of hot flushes of each severity you had during the night.

Complete this section in the evening	
Have you taken your study medication today?	Yes O No O
No hot flushes during the day	O
Total **number** of hot flushes of **each severity** during the day:	
MILD: *Sensation of heat without sweating*	⌊⌋⌊⌋ Mild
MODERATE: *Sensation of heat with sweating, but able to continue activity.*	⌊⌋⌊⌋ Moderate
SEVERE: *Sensation of heat with sweating, causing cessation (stopping) of activity.*	⌊⌋⌊⌋ Severe

Complete this section in the morning	
Total **number** of times you woke up last night:	⌊⌋⌊⌋ "0" if none
No hot flushes during the night	O
Total **number** of hot flushes of **each severity** during the night:	
MILD: *Sensation of heat without sweating*	⌊⌋⌊⌋ Mild
MODERATE: *Sensation of heat with sweating, but able to continue activity.*	⌊⌋⌊⌋ Moderate
SEVERE: *Sensation of heat with sweating, causing cessation (stopping) of activity.* **Note**: *A night-time hot flush that wakes you up (stops you sleeping) is a severe hot flush. If you had more hot flushes once you were awake, record these too.*	⌊⌋⌊⌋ Severe

Menopause-specific Quality-of-Life Questionnaire Intervention Version

THE MENOPAUSE-SPECIFIC

QUALITY-OF-LIFE QUESTIONNAIRE

INTERVENTION VERSION

MENQOL-I™

Primary Care Research Unit
Department of Family and Community Medicine
Sunnybrook Health Sciences Centre
University of Toronto

Authors: John R. Hilditch, Jacqueline E. Lewis, Peter G. Norton, Earl Dunn

The development of the MENQOL™ questionnaire was funded by CIBA-GEIGY Canada Ltd., Mississauga, Canada
The development of the MENQOL-I™ was funded by JANSSEN-ORTHO Inc., Canada

The authors request acknowledgement in any research publications in which the questionnaire is used.

For information on, or permission to use the questionnaire, please contact Mapi Research Trust,
E-mail: PROinformation@mapi-trust.org – Internet: www.proqolid.org

INSTRUCTIONS

Each of the items in the questionnaire is in the form of the examples below:

		Not at all bothered → Extremely bothered 0 1 2 3 4 5 6
NIGHT SWEATS	☐ No	☐ Yes → ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

Indicate whether or not you have experienced this problem in the **PAST MONTH**.

IF YOU ***HAVE NOT*** EXPERIENCED THE PROBLEM:

Mark "No"

NIGHT SWEATS	☑ No	☐ Yes → ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

→ Go to the next item.

IF YOU ***HAVE*** EXPERIENCED THE PROBLEM:

Mark "Yes", then check off how bothered you were by the problem.

NIGHT SWEATS	☐ No	☑ Yes → ☐ 0 ☐ 1 ☑ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

→ Go to the next item.

This questionnaire is completely confidential. Your name will not be associated with your responses. However, if for any reason you do not wish to complete an item, please leave it and go on to the next one.

The Menopause-Specific Quality of Life Questionnaire – Intervention Version *Page 3 of 4*

Date : _____/ _____/ _____
yy mm dd

Subject ID # : _____

For each of the following items, indicate whether you have experienced the problem in the **PAST MONTH**. If you have, rate how much you have been *bothered* by the problem.

			Not at all bothered → Extremely bothered						
			0	1	2	3	4	5	6
1. HOT FLUSHES OR FLASHES	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
2. NIGHT SWEATS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
3. SWEATING	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
4. DISSATISFACTION WITH MY PERSONAL LIFE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
5. FEELING ANXIOUS OR NERVOUS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
6. POOR MEMORY	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
7. ACCOMPLISHING LESS THAN I USED TO	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
8. FEELING DEPRESSED, DOWN OR BLUE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
9. BEING IMPATIENT WITH OTHER PEOPLE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
10. FEELINGS OF WANTING TO BE ALONE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
11. FLATULENCE (WIND) OR GAS PAINS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
12. ACHING IN MUSCLES AND JOINTS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
13. FEELING TIRED OR WORN OUT	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
14. DIFFICULTY SLEEPING	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
15. ACHES IN BACK OF NECK OR HEAD	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
16. DECREASE IN PHYSICAL STRENGTH	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

The Menopause-Specific Quality of Life Questionnaire – Intervention Version *Page 4 of 4*

Date :_____/ _____/ _____
yy mm dd

Subject ID # :_____

			Not at all bothered → Extremely bothered						
			0	1	2	3	4	5	6
17. DECREASE IN STAMINA	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
18. LACK OF ENERGY	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
19. DRY SKIN	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
20. WEIGHT GAIN	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
21. INCREASED FACIAL HAIR	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
22. CHANGES IN APPEAR-ANCE, TEXTURE OR TONE OF MY SKIN	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
23. FEELING BLOATED	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
24. LOW BACKACHE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
25. FREQUENT URINATION	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
26. INVOLUNTARY URINATION WHEN LAUGHING OR COUGHING	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
27. DECREASE IN MY SEXUAL DESIRE	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
28. VAGINAL DRYNESS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
29. AVOIDING INTIMACY	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
30. BREAST PAIN OR TENDERNESS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
31. VAGINAL BLEEDING OR SPOTTING	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
32. LEG PAINS OR CRAMPS	☐ No	☐ Yes →	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

INSTRUCTIONS FOR USE AND SCORING OF THE MENOPAUSE-SPECIFIC QUALITY OF LIFE QUESTIONNAIRE – INTERVENTION VERSION MENQOL-I™

USE:

1. The title page, subject questionnaire instruction and 32 items constitute the MENQOL-I™.

2. The MENQOL-I™ questionnaire is designed to be self-administered either in person or by mail. Use of electronic, verbal, Braille, sign language or other delivery methods require pre-testing.

3. The questionnaire requires, on average, 7 minutes to complete with a range of 5 to 15 minutes based on the original English and French Canadian pre-tests.

4. The official Canadian English-language MENQOL™ requires a grade 6 reading proficiency in a Canadian population.

5. Ensure you have the correct questionnaire recall period based upon your study need.

6. Questions 30, 31 and 32 have been added to the original MENQOL™ questionnaire for use in perimenopausal women or women taking hormone therapy or SERMS in an intervention trial.

REFERENCES :

Hilditch JR, Lewis JE, Ross AH, et al. A menopause-specific quality of life questionnaire: Development and psychometric properties. Maturitas 1996; 24:161-175.

Lewis JE, Hilditch JR, Wong CJ. Further psychometric property development of the Menopause-Specific Quality of Life questionnaire and development of a modified version, the MENQOL-Intervention questionnaire. Maturitas 2005; 50:209-221.

SCORING:

1. a) Each domain is scored separately.

   b) The scale contains four domains :

   i Vasomotor - Items 1 to 3
   ii Psychosocial - Items 4 to 10
   iii Physical - Items 11 to 26, 30, 31, 32
   iv Sexual - Items 27 to 29

2. For analyses, 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

3. Each domain mean ranges from 1 to 8. The overall questionnaire score is the mean of the domain means.

Hot Flash Related Daily Interference Scale (HFRDIS)

Author: Janet S. Carpenter, PhD, RN, Indiana University School of Nursing, Indiana, USA

HOT FLASH RELATED DAILY INTERFERENCE SCALE (HFRDIS)

DATE: _____

Circle the number that best describes how much hot flashes have interfered with each aspect of your life *during the past* week.

	Not at all										Very much so
1. Work (work outside the home and house work)	0	1	2	3	4	5	6	7	8	9	10
2. Social activities (time spent with family, friends, etc.)	0	1	2	3	4	5	6	7	8	9	10
3. Leisure activities (time spent relaxing, doing hobbies, etc.)	0	1	2	3	4	5	6	7	8	9	10
4. Sleep	0	1	2	3	4	5	6	7	8	9	10
5. Mood	0	1	2	3	4	5	6	7	8	9	10
6. Concentration	0	1	2	3	4	5	6	7	8	9	10
7. Relations with others	0	1	2	3	4	5	6	7	8	9	10
8. Sexuality	0	1	2	3	4	5	6	7	8	9	10
9. Enjoyment of life	0	1	2	3	4	5	6	7	8	9	10
10. Overall quality of life	0	1	2	3	4	5	6	7	8	9	10

TOTAL SCORE: _____

("Hot flashes interfered with my life": 0 = not at all; 100 = very much so)

Pittsburgh Sleep Quality Index (PSQI)

Authors: Buysse DJ; Berman SR; Kupfer DJ; Monk TH; Reynolds CF
University of Pittsburgh School of Medicine, USA.

PITTSBURGH SLEEP QUALITY INDEX

INSTRUCTIONS:
The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions.

1. During the past month, what time have you usually gone to bed at night?

   BED TIME _____

2. During the past month, how long (in minutes) has it usually taken you to fall asleep each night?

   NUMBER OF MINUTES _____

3. During the past month, what time have you usually gotten up in the morning?

   GETTING UP TIME _____

4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)

   HOURS OF SLEEP PER NIGHT _____

For each of the remaining questions, check the one best response. Please answer all questions.

5. During the past month, how often have you had trouble sleeping because you . . .

a) Cannot get to sleep within 30 minutes

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

b) Wake up in the middle of the night or early morning

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

c) Have to get up to use the bathroom

Not during the past month_____	Less than once a week_____	Once or twice a week_____	Three or more times a week_____

d) Cannot breathe comfortably

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

e) Cough or snore loudly

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

f) Feel too cold

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

g) Feel too hot

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

h) Had bad dreams

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

i) Have pain

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

j) Other reason(s), please describe_____

_____

How often during the past month have you had trouble sleeping because of this?

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

6. During the past month, how would you rate your sleep quality overall?

Very good _____

Fairly good _____

Fairly bad _____

Very bad _____

7. During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

8. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

No problem at all _____

Only a very slight problem _____

Somewhat of a problem _____

A very big problem _____

10. Do you have a bed partner or room mate?

No bed partner or room mate _____

Partner/room mate in other room _____

Partner in same room, but not same bed _____

Partner in same bed _____

If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .

a) Loud snoring

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

b) Long pauses between breaths while asleep

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

c) Legs twitching or jerking while you sleep

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

d) Episodes of disorientation or confusion during sleep

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

e) Other restlessness while you sleep; please describe_____

_____

Not during the past month_____ Less than once a week_____ Once or twice a week_____ Three or more times a week_____

Insomnia Severity Index (ISI)

Author: Charles M. Morin, Ph.D., Professor of Psychology, Université Laval, Quebec, Canada

Insomnia Severity Index

The Insomnia Severity Index has seven questions. The seven answers are added up to get a total score. When you have your total score, look at the 'Guidelines for Scoring/Interpretation' below to see where your sleep difficulty fits.

For each question, please CIRCLE the number that best describes your answer.

Please rate the CURRENT (i.e. LAST 2 WEEKS) SEVERITY of your insomnia problem(s).

Insomnia Problem	None	Mild	Moderate	Severe	Very Severe
1. Difficulty falling asleep	0	1	2	3	4
2. Difficulty staying asleep	0	1	2	3	4
3. Problems waking up too early	0	1	2	3	4

4. How SATISFIED/DISSATISFIED are you with your CURRENT sleep pattern?

Very Satisfied	Satisfied	Moderately Satisfied	Dissatisfied	Very Dissatisfied
0	1	2	3	4

5. How NOTICEABLE to others do you think your sleep problem is in terms of impairing the quality of your life?

Not at all Noticeable	A Little	Somewhat	Much	Very Much Noticeable
0	1	2	3	4

6. How WORRIED/DISTRESSED are you about your current sleep problem?

Not at all Worried	A Little	Somewhat	Much	Very Much Worried
0	1	2	3	4

7. To what extent do you consider your sleep problem to INTERFERE with your daily functioning (e.g. daytime fatigue, mood, ability to function at work/daily chores, concentration, memory, mood, etc.) CURRENTLY?

Not at all Interfering	A Little	Somewhat	Much	Very Much Interfering
0	1	2	3	4

Guidelines for Scoring/Interpretation:

Add the scores for all seven items (questions 1 + 2 + 3 + 4 + 5 +6 + 7) = _____ your total score

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

Beck Depression Inventory II

Authors: Aaron T. Beck, Robert A. Steer, Gregory K. Brown

Instructions: This questionnaire consists of 21 groups of statements. Please read each group of statements carefully, and then pick out the **one statement** in each group that best describes the way you have been feeling during the **past two weeks, including today**. Circle the number beside the statement you have picked. If several statements in the group seem to apply equally well, circle the highest number for that group. Be sure that you do not choose more than one statement for any group, including Item 16 (Changes in Sleeping Pattern) or Item 18 (Changes in Appetite).

1. Sadness

0 I do not feel sad.
1 I feel sad much of the time.
2 I am sad all the time.
3 I am so sad or unhappy that I can't stand it.

2. Pessimism

0 I am not discouraged about my future.
1 I feel more discouraged about my future than I used to be.
2 I do not expect things to work out for me.
3 I feel my future is hopeless and will only get worse.

3. Past Failure

0 I do not feel like a failure.
1 I have failed more than I should have.
2 As I look back, I see a lot of failures.
3 I feel I am a total failure as a person.

4. Loss of Pleasure

0 I get as much pleasure as I ever did from the things I enjoy.
1 I don't enjoy things as much as I used to.
2 I get very little pleasure from the things I used to enjoy.
3 I can't get any pleasure from the things I used to enjoy.

5. Guilty Feelings

0 I don't feel particularly guilty.
1 I feel guilty over many things I have done or should have done.
2 I feel quite guilty most of the time.
3 I feel guilty all of the time.

6. Punishment Feelings

0 I don't feel I am being punished.
1 I feel I may be punished.
2 I expect to be punished.
3 I feel I am being punished.

7. Self-Dislike

0 I feel the same about myself as ever.
1 I have lost confidence in myself.
2 I am disappointed in myself.
3 I dislike myself.

8. Self-Criticalness

0 I don't criticize or blame myself more than usual.
1 I am more critical of myself than I used to be.
2 I criticize myself for all of my faults.
3 I blame myself for everything bad that happens.

9. Suicidal Thoughts or Wishes

0 I don't have any thoughts of killing myself.
1 I have thoughts of killing myself, but I would not carry them out.
2 I would like to kill myself.
3 I would kill myself if I had the chance.

10. Crying

0 I don't cry anymore than I used to.
1 I cry more than I used to.
2 I cry over every little thing.
3 I feel like crying, but I can't.

11. Agitation

0 I am no more restless or wound up than usual.
1 I feel more restless or wound up than usual.
2 I am so restless or agitated that it's hard to stay still.
3 I am so restless or agitated that I have to keep moving or doing something.

12. Loss of Interest

0 I have not lost interest in other people or activities.
1 I am less interested in other people or things than before.
2 I have lost most of my interest in other people or things.
3 It's hard to get interested in anything.

13. Indecisiveness

0 I make decisions about as well as ever.
1 I find it more difficult to make decisions than usual.
2 I have much greater difficulty in making decisions than I used to.
3 I have trouble making any decisions.

14. Worthlessness

0 I do not feel I am worthless.
1 I don't consider myself as worthwhile and useful as I used to.
2 I feel more worthless as compared to other people.
3 I feel utterly worthless.

15. Loss of Energy

0 I have as much energy as ever.
1 I have less energy than I used to have.
2 I don't have enough energy to do very much.
3 I don't have enough energy to do anything.

16. Changes in Sleeping Pattern

0 I have not experienced any change in my sleeping pattern.
1a I sleep somewhat more than usual.
1b I sleep somewhat less than usual.
2a I sleep a lot more than usual.
2b I sleep a lot less than usual.
3a I sleep most of the day.
3b I wake up 1~2 hours early and can't get back to sleep.

17. Irritability

0 I am no more irritable than usual.
1 I am more irritable than usual.
2 I am much more irritable than usual.
3 I am irritable all the time.

18. Changes in Appetite

0 I have not experienced any change in my appetite.
1a My appetite is somewhat less than usual.
1b My appetite is somewhat greater than usual.
2a My appetite is much less than before.
2b My appetite is much greater than usual.
3a I have no appetite at all.
3b I crave food all the time.

19. Concentration Difficulty

0 I can concentrate as well as ever.
1 I can't concentrate as well as usual.
2 It's hard to keep my mind on anything for very long.
3 I find I can't concentrate on anything.

20. Tiredness or Fatigue

0 I am no more tired or fatigued than usual.
1 I get more tired or fatigued more easily than usual.
2 I am too tired or fatigued to do a lot of the things I used to do.
3 I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex

0 I have not noticed any recent change in my interest in sex.
1 I am less interested in sex than I used to be.
2 I am much less interested in sex now.
3 I have lost interest in sex completely.

Scoring the Beck Depression Inventory

After you have completed the questionnaire, add up the score for each of the 21 questions. The following table indicates the relationship between total score and level of depression according to the Beck Depression Inventory.

Classification	Total Score	Level of Depression
Low	1-10	Normal ups and downs
	11-16	Mild mood disturbance
Moderate	17-20	Borderline clinical depression
	21-30	Moderate depression
Significant	31-40	Severe depression
	Over 40	Extreme depression

Electronic Columbia Suicide Severity Rating Scale (eC-SSRS)

Below is an example of the C-SSRS. In this study, an electronic version of the C-SSRS will be used in which subjects self-report using an IVR system.

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

Posner, Brent, Lucas, Gould, Stanley, Brown, Fisher, Zelazny, Burke, Oquendo, & Mann

RISK ASSESSMENT VERSION
(* elements added with permission for Lifeline centers)

Instructions: Check all risk and protective factors that apply. To be completed following the patient interview, review of medical record(s) and/or consultation with family members and/or other professionals.

Suicidal and Self-Injury Behavior (Past week)		**Clinical Status (Recent)**	
☐	Actual suicide attempt ☐ Lifetime	☐	Hopelessness
☐	Interrupted attempt ☐ Lifetime	☐	Helplessness*
☐	Aborted attempt ☐ Lifetime	☐	Feeling Trapped*
☐	Other preparatory acts to kill self ☐ Lifetime	☐	Major depressive episode
☐	Self-injury behavior w/o suicide intent ☐ Lifetime	☐	Mixed affective episode
Suicide Ideation (Most Severe in Past Week)		☐	Command hallucinations to hurt self
☐	Wish to be dead	☐	Highly impulsive behavior
☐	Suicidal thoughts	☐	Substance abuse or dependence
☐	Suicidal thoughts with method (but without specific plan or intent to act)	☐	Agitation or severe anxiety
☐	Suicidal intent (without specific plan)	☐	Perceived burden on family or others
☐	Suicidal intent with specific plan	☐	Chronic physical pain or other acute medical problem (AIDS, COPD, cancer, etc.)
Activating Events (Recent)		☐	Homicidal ideation
☐	Recent loss or other significant negative event	☐	Aggressive behavior towards others
	Describe:	☐	Method for suicide available (gun, pills, etc.)
		☐	Refuses or feels unable to agree to safety plan
☐	Pending incarceration or homelessness	☐	Sexual abuse (lifetime)
☐	Current or pending isolation or feeling alone	☐	Family history of suicide (lifetime)
Treatment History		**Protective Factors (Recent)**	
☐	Previous psychiatric diagnoses and treatments	☐	Identifies reasons for living
☐	Hopeless or dissatisfied with treatment	☐	Responsibility to family or others; living with family
☐	Noncompliant with treatment	☐	Supportive social network or family
☐	Not receiving treatment	☐	Fear of death or dying due to pain and suffering
Other Risk Factors		☐	Belief that suicide is immoral, high spirituality
☐		☐	Engaged in work or school
		☐	Engaged with Phone Worker *
		Other Protective Factors	
		☐	

Describe any suicidal, self-injury or aggressive behavior (include dates):

Lifeline Version 1/2014

SUICIDAL IDEATION		
Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.	**Lifetime: Time He/She Felt Most Suicidal**	**Past 1 month**
1. Wish to be Dead Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up. ***Have you wished you were dead or wished you could go to sleep and not wake up?*** If yes, describe:	Yes ☐ No ☐	Yes ☐ No ☐
2. Non-Specific Active Suicidal Thoughts General non-specific thoughts of wanting to end one's life/commit suicide (e.g., *"I've thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period. ***Have you actually had any thoughts of killing yourself?*** If yes, describe:	Yes ☐ No ☐	Yes ☐ No ☐
3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act Subject endorses thoughts of suicide and has thought of at least one method during the assessment period. This is different than a specific plan with time, place or method details worked out (e.g., thought of method to kill self but not a specific plan). Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it...and I would never go through with it."* ***Have you been thinking about how you might do this?*** If yes, describe:	Yes ☐ No ☐	Yes ☐ No ☐
4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan Active suicidal thoughts of killing oneself and subject reports having some intent to act on such thoughts, as opposed to *"I have the thoughts but I definitely will not do anything about them."* ***Have you had these thoughts and had some intention of acting on them?*** If yes, describe:	Yes ☐ No ☐	Yes ☐ No ☐
5. Active Suicidal Ideation with Specific Plan and Intent Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out. ***Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?*** If yes, describe:	Yes ☐ No ☐	Yes ☐ No ☐
INTENSITY OF IDEATION		
The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe). Ask about time he/she was feeling the most suicidal. Lifetime - *Most Severe Ideation:* _____ (*Type # (1-5)*) _____ (*Description of Ideation*) Recent - *Most Severe Ideation:* _____ (*Type # (1-5)*) _____ (*Description of Ideation*)	Most Severe	Most Severe
Frequency ***How many times have you had these thoughts?*** (1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day	____	____
Duration ***When you have the thoughts how long do they last?*** (1) Fleeting - few seconds or minutes (2) Less than 1 hour/some of the time (3) 1-4 hours/a lot of time (4) 4-8 hours/most of day (5) More than 8 hours/persistent or continuous	____	____
Controllability ***Could/can you stop thinking about killing yourself or wanting to die if you want to?*** (1) Easily able to control thoughts (2) Can control thoughts with little difficulty (3) Can control thoughts with some difficulty (4) Can control thoughts with a lot of difficulty (5) Unable to control thoughts (0) Does not attempt to control thoughts	____	____
Deterrents ***Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?*** (1) Deterrents definitely stopped you from attempting suicide (2) Deterrents probably stopped you (3) Uncertain that deterrents stopped you (4) Deterrents most likely did not stop you (5) Deterrents definitely did not stop you (0) Does not apply	____	____
Reasons for Ideation ***What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?*** (1) Completely to get attention, revenge or a reaction from others (2) Mostly to get attention, revenge or a reaction from others (3) Equally to get attention, revenge or a reaction from others and to end/stop the pain (4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (0) Does not apply	____	____

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Lifetime**	**Past 3 months**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself. Intent does not have to be 100%. If there is *any* intent/desire to die associated with the act, then it can be considered an actual suicide attempt. *There does not have to be any injury or harm*, just the potential for injury or harm. If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt. Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances. For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e.g., gunshot to head, jumping from window of a high floor/story). Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred. ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you _____?*** ***Or Did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** (Self-Injurious Behavior without suicidal intent) If yes, describe:	Yes ☐ No ☐ Total # of Attempts _____	Yes ☐ No ☐ Total # of Attempts _____
Has subject engaged in Non-Suicidal Self-Injurious Behavior?	Yes ☐ No ☐	Yes ☐ No ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting. Once they ingest any pills, this becomes an attempt rather than an interrupted attempt. Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger. Once they pull the trigger, even if the gun fails to fire, it is an attempt. Jumping: Person is poised to jump, is grabbed and taken down from ledge. Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so. ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of interrupted _____	Yes ☐ No ☐ Total # of interrupted _____
Aborted or Self-Interrupted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior. Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else. ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of aborted or self-interrupted _____	Yes ☐ No ☐ Total # of aborted or self-interrupted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt. This can include anything beyond a verbalization or thought, such as assembling a specific method (e.g., buying pills, purchasing a gun) or preparing for one's death by suicide (e.g., giving things away, writing a suicide note). ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	Yes ☐ No ☐ Total # of preparatory acts _____	Yes ☐ No ☐ Total # of preparatory acts _____

	Most Recent Attempt Date:	Most Lethal Attempt Date:	Initial/First Attempt Date:
Actual Lethality/Medical Damage: 0. No physical damage or very minor physical damage (e.g., surface scratches). 1. Minor physical damage (e.g., lethargic speech; first-degree burns; mild bleeding; sprains). 2. Moderate physical damage; medical attention needed (e.g., conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel). 3. Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e.g., comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures). 4. Severe physical damage; *medical* hospitalization with intensive care required (e.g., comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area). 5. Death	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over). 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____

Appendix 16.1.9 Documentation of Statistical Methods

- Statistical Analysis Plans
  - 814-PM-02 Statistical Analysis Plan (Final Version 3) 11 December 2019
  - 814-PM-02 Statistical Analysis Plan (Final Version 2) 18 June 2019
  - 814-PM-02 Statistical Analysis Plan (Final Version 1) 20 December 2018

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: **(Version Date)**	814-PM-02 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, SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX United Kingdom
Sponsor Representative:	Stephen Pawsey MBBS FRCA FFPM NeRRe Therapeutics Ltd SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX, UK Phone: +44 (0) 7827 460726 Email: steve.pawsey@nerretherapeutics.com
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:
Christelle Pommie
Cytel Inc. Geneva Branch|
20 Route de Pre-Bois
CH-1216 Cointrin

Signature: Christelle Pommie *Electronically signed by: Christelle Pommie 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:
Stephen Pawsey
NeRRe Therapeutics Ltd
SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage, SG1 2FX, UK

Signature: S.Pawsey *Electronically signed by: S.Pawsey 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** IF Q4 ≥ 7, THEN set value to 0 IF Q4 < 7 and ≥ 6, THEN set value to 1 IF Q4 < 6 and ≥ 5, THEN set value to 2 IF Q4 < 5, THEN set value to 3 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** Q6 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQIMEDS	**NEED MEDS TO SLEEP** Q7 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQI	**TOTAL** DURAT + DISTB + LATEN + DAYDYS + HSE + SLPQUAL + MEDS Minimum Score = 0 (better); Maximum Score = 21 (worse) Interpretation: TOTAL ≤ 5 associated with good sleep quality 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) Note: number of capsules planned to be received is 4*duration of dosing (in days). 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. 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: In Placebo group, Cumulative dose = 0 In NT-814 40mg group, Cumulative dose = (number of capsules taken / 4) * 40 mg. In NT-814 80mg group, Cumulative dose = (number of capsules taken / 4) * 80 mg. In NT-814 120mg group, Cumulative dose = (number of capsules taken / 4) * 120 mg. 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)

Cytel	

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: **(Version Date)**	814-PM-02 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, SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX United Kingdom
Sponsor Representative:	Stephen Pawsey MBBS FRCA FFPM NeRRe Therapeutics Ltd SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX, UK Phone: +44 (0) 7827 460726 Email: steve.pawsey@nerretherapeutics.com
Document Date:	18 June 2019
Document Version:	Final Version 2.0

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:	18 June 2019 / Final Version 2.0

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

Signature: _____
Date: 18-JUN-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:
Stephen Pawsey
NeRRe Therapeutics Ltd
SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage, SG1 2FX, UK

Signature: _____
Date: 18 JUNE 2019.

TABLE OF CONTENTS

Section **Page**

LIST OF IN-TEXT TABLES

Table **Page**

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 4.0 dated 13 May 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 used treatments that may have had a confounding effect on hot flash frequency and/or severity within the 28 days before screening. The medications entered in EDC will be reviewed by Sponsor's blinded Medical Expert, and a final list of ATC classification (Level 2 or other) will be confirmed during the data review meeting prior to DB lock. 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 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 will be finalized during the blinded data review prior to database lock and unblinding of treatment allocation.

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, the occurrence date and the categorization as 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, at a blinded data review meeting (BDRM).

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 severity = Sum of [(frequency x severity) / number of hot flushes] 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). If the number of hot flushes is equal to 0, then the mean severity will be set to 0.

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.

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	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.

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 and mean severity 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. These 7 last days with available data should be within the 14 days prior to the study visit, so in the following window: [Visit date – 14; Visit date], with at least one day within the last 7 days prior to the study visit. 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) On frequency of hot flushes:

- 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) On severity of hot flushes:

- 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 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 average hot flush severity 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 and early termination 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 additional sensitivity 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 section4.3.2.

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 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 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** IF Q4 ≥ 7, THEN set value to 0 IF Q4 < 7 and ≥ 6, THEN set value to 1 IF Q4 < 6 and ≥ 5, THEN set value to 2 IF Q4 < 5, THEN set value to 3 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** Q6 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQIMEDS	**NEED MEDS TO SLEEP** Q7 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQI	**TOTAL** DURAT + DISTB + LATEN + DAYDYS + HSE + SLPQUAL + MEDS Minimum Score = 0 (better); Maximum Score = 21 (worse) Interpretation: TOTAL ≤ 5 associated with good sleep quality 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. However, if at the time of the BDRM an excessive amount of missing data are observed on the questionnaires, imputation rules will be put in place. Any such rules will be documented in an amendment of the SAP.

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) Note: number of capsules planned to be received is 4*duration of dosing (in days). 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. In the event that the subject doesn't return some or all of the 4 weekly cards, 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: In Placebo group, Cumulative dose = 0 In NT-814 40mg group, Cumulative dose = (number of capsules taken / 4) * 40 mg. In NT-814 80mg group, Cumulative dose = (number of capsules taken / 4) * 80 mg. In NT-814 120mg group, Cumulative dose = (number of capsules

	taken / 4) * 120 mg. 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.

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.

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).

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.

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)
Category 10 – Completed Suicide

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 6-10) 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 1-10) 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 10 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 score. Counts 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 14.1.2	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 14.2.1.2C	Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline - MMRM Analysis (Modified 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 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.

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 16.2.6.3.2 Insomnia Severity Index Score (Safety Analysis Set)

Listing 16.2.6.4 HFRDIS Scores (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	Urine Pregnancy Test (Safety Analysis Set)

Other Safety Data (CSR Appendix 16.2.9)

Listing 16.2.9.1	Vital Signs (Safety Analysis Set)
Listing 16.2.9.2	Physical Examination (Safety Analysis Set)
Listing 16.2.9.3	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	Plot of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (Full Analysis Set)
Figure 14.2.1.1B	Plot of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (PP Analysis Set)
Figure 14.2.1.2A	Bar Graph of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group (Full Analysis Set)
Figure 14.2.1.2B	Bar Graph of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group (PP Analysis Set)
Figure 14.2.2.1A	Plot of Mean Daily Severity of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (Full Analysis Set)
Figure 14.2.2.1B	Plot of Mean Daily Severity of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (PP Analysis Set)
Figure 14.2.2.2A	Bar Graph of Mean Daily Severity of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group (Full Analysis Set)
Figure 14.2.2.2B	Bar Graph of Mean Daily Severity of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group (PP Analysis Set)
Figure 14.2.3	Plot of Mean Daily Frequency of All Hot Flushes Change From Baseline over Time by Treatment Group (Full Analysis Set)
Figure 14.2.4	Plot of Mean Daily Severity of All Hot Flushes Change From Baseline over Time by Treatment Group (Full Analysis Set)

Cytel	

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: **(Version Date)**	814-PM-02 Version 1 dated 09 August 2018
Name of Test Drug:	NT-814
Phase:	2b
Methodology:	Randomized, Double-blind, placebo controlled, adaptive
Sponsor:	NeRRe Therapeutics Ltd, SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX United Kingdom
Sponsor Representative:	Stephen Pawsey MBBS FRCA FFPM NeRRe Therapeutics Ltd SBC, Incubator Building, Gunnels Wood Rd., Stevenage, SG1 2FX, UK Phone: +44 (0) 7827 460726 Email: steve.pawsey@nerretherapeutics.com
Document Date:	20 December 2018
Document Version:	Final Version 1.0

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:	20 December 2018 / Final Version 1.0

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

Signature: _____
Date: 21 DEC 2018

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:
Stephen Pawsey
NeRRe Therapeutics Ltd
SBC, Incubator Building,
Gunnels Wood Rd.,
Stevenage, SG1 2FX, UK

Signature: _____
Date: 21 DEC 2018

TABLE OF CONTENTS

Section **Page**

LIST OF IN-TEXT TABLES

Table **Page**

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 1.0, 9 Aug 2018, and to the annotated CRF Final Version dated 30 October 2018.

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
Biochemistry and Haematology		X	X	X	X		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.

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 (haematology, clinical chemistry, urinalysis) at Weeks 2, 4, 12 and 16
- 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.

Per Protocol (PP) Set: All subjects in the FAS 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 will be finalized during the blinded data review prior to database lock and unblinding of treatment allocation.

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, the occurrence date and the categorization as 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, at a blinded data review meeting (BDRM).

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.

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 severity = Sum of [(frequency x severity) / number of hot flushes] 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).

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.

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	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.

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 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 and mean severity 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. These 7 last days with available data should be within the 14 days prior to the study visit, so in the following window: [Visit date – 14; Visit date], with at least one day within the last 7 days prior to the study visit. 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) On severity of hot flushes:

- 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. 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 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 average hot flush severity will be summarised by treatment group.

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 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 and early termination 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 additional sensitivity 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 section4.3.2.

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 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 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
IF Q4 ≥ 7, THEN set value to 0
IF Q4 < 7 and ≥ 6, THEN set value to 1
IF Q4 < 6 and ≥ 5, THEN set value to 2
IF Q4 < 5, THEN set value to 3
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** Q6 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQIMEDS	**NEED MEDS TO SLEEP** Q7 Minimum Score = 0 (better); Maximum Score = 3 (worse)
PSQI	**TOTAL** DURAT + DISTB + LATEN + DAYDYS + HSE + SLPQUAL + MEDS Minimum Score = 0 (better); Maximum Score = 21 (worse) Interpretation: TOTAL ≤ 5 associated with good sleep quality 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. However, if at the time of the BDRM an excessive amount of missing data are observed on the questionnaires, imputation rules will be put in place. Any such rules will be documented in an amendment of the SAP.

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) Note: number of capsules planned to be received is 4*duration of dosing (in days). 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. In the event that the subject doesn't return some or all of the 4 weekly cards, 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: In Placebo group, Cumulative dose = 0 In NT-814 40mg group, Cumulative dose = (number of capsules taken / 4) * 40 mg. In NT-814 80mg group, Cumulative dose = (number of capsules taken / 4) * 80 mg. In NT-814 120mg group, Cumulative dose = (number of capsules

	taken / 4) * 120 mg. 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.

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.

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
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.

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).

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.

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)
Category 10 – Completed Suicide

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 6-10) 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 1-10) 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 10 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 score. Counts 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 14.1.2	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 14.2.1.2A	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 14.2.1.3A	Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline - ANCOVA Analysis (Full Analysis Set)
Table 14.2.1.3B	Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline - ANCOVA Analysis (PP 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.

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 16.2.6.3.2	Insomnia Severity Index Score (Safety Analysis Set)
Listing 16.2.6.4	HFRDIS Scores (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	Urine Pregnancy Test (Safety Analysis Set)

Other Safety Data (CSR Appendix 16.2.9)

Listing 16.2.9.1	Vital Signs (Safety Analysis Set)
Listing 16.2.9.2	Physical Examination (Safety Analysis Set)
Listing 16.2.9.3	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	Plot of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (Full Analysis Set)
Figure 14.2.1.1B	Plot of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline over Time by Treatment Group (PP Analysis Set)
Figure 14.2.1.2A	Bar Graph of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group (Full Analysis Set)
Figure 14.2.1.2B	Bar Graph of Mean Daily Frequency of Moderate and Severe Hot Flushes Change From Baseline at Week 4 and 12 by Treatment Group