[redacted]

NeRRe

VOLCANO-2

NCT02993822

A Double-Blind, Randomized, Placebo Controlled Study of the Efficacy and Safety of Three Doses of Orvepitant in Subjects with Chronic Refractory Cough

Statistical Analysis Plan

08 May 2018

[redacted]

SPONSOR SIGNATURE PAGE

Approved by: D. Jolliffe

Darren Jolliffe
Statistician
NeRRe

Date 08MAY2018

Approved by:

Steve Pawsey
Chief Medical Officer
NeRRe

08 MAY 2018
Date

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACM	Automated cough monitor
AE	Adverse Event
ALT	Alanine aminotransferase
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
BLLQ	Below lower limit of quantification
CRC	Chronic refractory cough
CQLQ	Cough-specific quality-of life questionnaire
CV	Coefficient of variation
DAP	Data analysis plan
eCRF	Electronic case report form
ECG	Electrocardiogram
ER	Exposure response
FAS	Full analysis set
GCP	Good clinical practice
GGT	Gamma glutamyl transferase
IMP	Investigational medicinal product
LCQ	Leicester cough questionnaire
LLoQ	Lower limit of quantification
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed model for repeated measures
PD	Pharmacodynamic
PK	Pharmacokinetic
Pop PK	Population Pharmacokinetic
PP	Per protocol
PT	Preferred term
QD	Once daily
RBC	Red blood cell
SAP	Statistical analysis plan
SD	Standard deviation
SOC	System organ class
TEAE	Treatment-emergent adverse event
VAS	Visual analogue scale
WBC	White blood cell

1 INTRODUCTION

To date, five phase 1 studies have been completed in healthy human subjects with orvepitant. There have also been five Phase 2 studies completed with orvepitant utilizing doses up to 60 mg once daily (QD).

In the 'VOLCANO-1' study with orvepitant, which was an open-label trial undertaken in 13 subjects with chronic refractory cough (CRC), the same target population to that in the current study, for which the mean cough history duration of this population was 13.1 years with a median 11.7 years, a statistically significant reduction was observed for the primary endpoint ($p<0.001$) of change in objectively recorded daytime cough frequency compared to Baseline following four weeks of treatment with orvepitant (30 mg, QD). Statistically significant improvements were also seen for all secondary objective and subjective measures of cough frequency and severity as well as in the Cough-specific Quality-of-Life Questionnaire (CQLQ) and Global Rating of Change for Cough Frequency and Severity assessments.

Thus far, the data supports the notion that NK-1 receptor antagonists generally, and more specifically orvepitant with its combined peripheral and central mechanisms of action, may prove of benefit in the treatment of CRC. The anticipated benefits of orvepitant as a centrally penetrant, potent and selective NK-1 receptor antagonist include rapid and sustained control of cough together with associated enhanced general Quality-of-Life.

The overall risk profile of orvepitant has been shown to be acceptable in the studies completed to date and the unmet need in this population and lack of treatment options supports the development of orvepitant for the CRC indication.

The objective of this study is to evaluate the efficacy and safety of orvepitant in adult subjects suffering from CRC and to identify a suitable dose of this NK-1 antagonist with which to initiate a pivotal Phase 3 registration program.

The scope of this statistical analysis plan (SAP) is to describe in detail the statistical analyses described in section 11 of the protocol. As stated in section 11.7 of the protocol, the plasma orvepitant concentrations will be analysed using a population pharmacokinetic (Pop PK) approach by nonlinear mixed effects modelling (NONMEM®). The details of this analysis will be described in a separate Pop PK Data Analysis Plan (DAP).

An electronic Case Report Form (eCRF) will be used to capture subject data into a secure, validated database. Automated cough monitor (ACM), pharmacokinetic (PK) and safety laboratory data will be transferred electronically into the database periodically during the study.

This SAP is based upon the following study documents:

- Study Protocol, Version 5.0 (Feb 21, 2018)
- eCRF, Version 4.0 (Sep 22, 2017)

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.

2 STUDY OBJECTIVES

2.1 Primary objectives

- To evaluate the efficacy of once daily doses of 10 mg, 20 mg, and 30 mg orvepitant versus placebo in reducing awake objective cough frequency.

2.2 Secondary objectives

- To evaluate the efficacy of once daily dosing of 10 mg, 20 mg, and 30 mg orvepitant versus placebo in:
  - Reducing 24-hour objective cough frequency
  - Reducing night-time (non-waking time) objective cough frequency
  - Reducing subject assessed cough severity and Urge-to-Cough
  - Improving Quality-of-Life and subject perception of change
- To evaluate the dose-response relationship of 10 mg, 20 mg, and 30 mg orvepitant
- To assess the safety and tolerability of 10 mg, 20 mg, and 30 mg orvepitant versus placebo over 12 weeks dosing

2.3 Pharmacokinetic objectives

- To evaluate the exposure-response relationship of orvepitant using population PK with sparse sampling.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

This is a phase II, multi-centre, double blind, randomised, parallel group, placebo-controlled dose range study in subjects with CRC. The primary endpoint is change from Baseline to Week 12 in awake objective cough frequency measured with an ACM, the 'Vitalojak™'. The aim of the comparison is to show superiority of orvepitant compared to placebo.

Subjects will be randomised in a 1:1:1:1 ratio to four parallel groups, 10 mg once per day orvepitant, 20 mg once per day orvepitant, 30 mg once per day orvepitant, and placebo

once per day. Randomisation will be stratified by region, namely North America and Europe.

Both orvepitant (10 mg, 20 mg, and 30 mg) and placebo will be presented as white tablets, identical in size and shape.

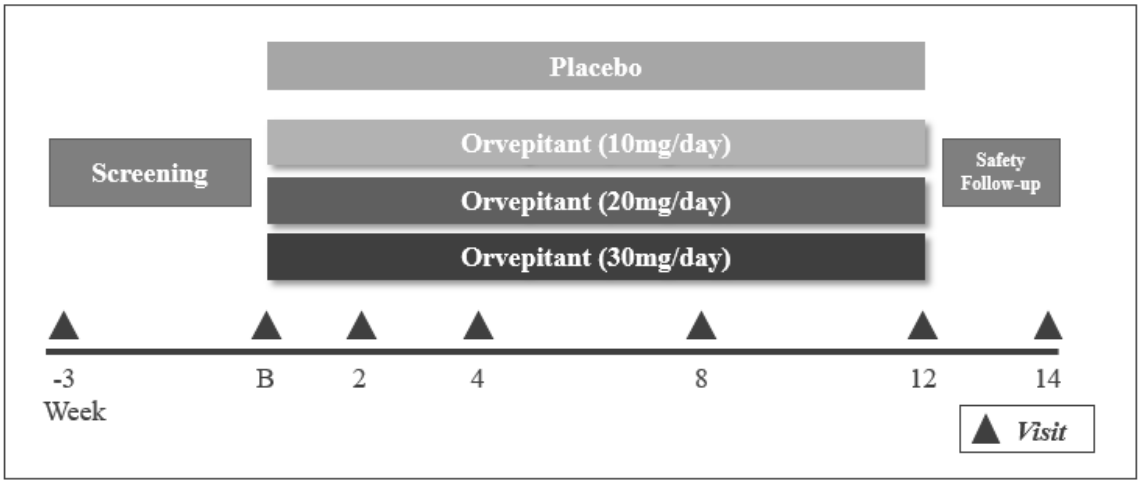
It is anticipated that approximately 292 subjects will be enrolled to ensure approximately 260 completers (~65 subjects per group; assuming approximately 10% drop out rate). The number of subjects enrolled may, however, be increased to a maximum of 320 in the event that the interim sample size re-estimation indicates that 292 is insufficient (see Section 4.10).

The planned duration of participation in the study for an individual subject is up to 18 weeks, which includes the following:

- Initial Screening phase of up to three weeks (can be extended to up to four weeks in the event that the ACM fails and the 24-hour cough monitoring must be repeated)
- 12-week double blind treatment phase
- Two-week safety follow-up

See **Figure 1** for study design schematic.

Figure 1: Study schematic



The assessments to be performed at each study visit are detailed in **Table 1: Schedule of Events**.

Table 1: Schedule of Events

	Screening Day -21 to 1[j]	**Baseline / Day 1**	**Wk 2 (± 3 days)[k]**	**Wk 4 (± 3 days)[k]**	**Wk 8 (± 5 days)[k]**	**Wk 12 (± 5 days)[k]**	**Early Termination**	**Wk 14 follow-up (± 5 days[k])**
Informed Consent	X							
Medical History	X							
Physical Examination[i]	X	X[c]					X	X
Spirometry	X							
Incl/Excl Criteria	X	X[c]						
Randomization		X						
Vital Signs[a]	X	X[c]	X	X	X	X	X	X
12-lead ECG	X	X[c]	X	X	X	X	X	X
Ambulatory cough monitoring[b]	X[g]		X	X		X		
LCQ		X[c]	X	X	X	X		
Global Rating of Change			X	X	X	X		
Cough Severity VAS		X[c]	X	X	X	X		
Urge-to-Cough VAS		X[c]	X	X	X	X		
PK sample			X[h]	X[h]	X[h]	X[h]		
Safety labs[e]	X	X[c]		X[h]	X[h]	X[h]	X	X
Opiate drug screen		X[c]	X	X	X	X		
Pregnancy test[f]	X	X[c]		X[g]	X[g]	X[g]	X	X
Dosing[d]		X----------X						
AE recording	X----------X							
Con Med recording	X----------X							

a Vital signs: systolic and diastolic blood pressure, pulse, temperature; weight (weight at Baseline/Day 1, Week 12 and Follow-up only)
b ACM (Vitalojak™) to be fitted during the visit and initiated prior to the subject leaving the clinic, monitoring will be continuous for 24hrs and device switched off after this period. Subjects will either return the ACM to the clinic (during Screening this must be done the following day) or collection from the subjects will be organized
c Performed prior to start of dosing
d Subjects are given their first dose of IMP in the clinic. All doses to be taken at approximately the same time each day; on clinic visits subjects will take their dose in the clinic after the pre-dose PK and safety samples, as applicable, have been taken. All other assessments can be performed post-dose
e Safety labs include blood samples and urine samples
f Pregnancy test in women of child-bearing potential only; serum pregnancy test at Screening, urine pregnancy at Baseline/Day 1, Wks 4, 8, and 12
g If the ACM fails during the Screening assessment and no useable recording is generated, the 24hr Screening ACM may be repeated and in this circumstance the Screening period may be extended by up to 1 week (to a maximum of 28 days)

h All blood samples are taken immediately prior to the dose.
i Full physical examination at Screening; symptom directed physical examination at subsequent visits
j Screening period may be extended for up to 1 week should the original ACM fail and this need to be repeated
k All visits and visit windows relate to Day 1
g Urine samples are taken immediately prior to the dose.

3.2 Efficacy and Safety Variables

3.2.1 Efficacy Assessments

- Subjects will be fitted with an ACM (the Vitalojak™) to record objective cough frequency over 24 hours at Screening, Weeks 2, 4, and 12
- Subjects will complete the Leicester Cough Questionnaire (LCQ) at the clinic at Baseline/Day 1 and Weeks 2 ,4 ,8, and 12
- Subjects will complete the Global Rating of Change for Cough Frequency & Severity Scale at the clinic at Weeks 2, 4, 8, and 12
- Subjects will complete both the Cough Severity Visual Analogue Scale (VAS) and Urge-to-Cough VAS at the clinic at Baseline/Day 1 and Weeks 2, 4, 8, and 12

3.2.2 Safety Assessments

Safety and tolerability will be assessed by the following:

- Physical examinations
- Change from Baseline in electrocardiogram (ECG) variables (heart rate and rhythm and RR, PR, QRS, QT, QTcF and QTcB intervals)
- Change from Baseline in clinical laboratory assessments (haematology, clinical chemistry, urinalysis)
- Change from Baseline in vital signs
- Adverse events
- Use of concomitant medications

3.2.3 Pharmacokinetics

Plasma levels of orvepitant will be measured by analysis of samples. Blood samples for analysis of plasma orvepitant concentrations will be collected at the clinic visits at Weeks 2, 4, 8 and 12, immediately prior (<15 mins) to the dose to be taken in the clinic.

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

'Baseline' is defined as the last available pre-treatment assessment. If time is missing for an assessment on the same day as first dose of treatment, then we shall assume the assessment to be pre-treatment. 'End of Study' is defined as the last available post-

treatment assessment. 'Study Day' will be calculated relative to the date of treatment start i.e. if Assessment Date < Treatment Start Date then;

$$\text{Study Day} = \text{Assessment Date} - \text{Treatment Start Date}$$

Else if Assessment Date ≥ Treatment Start Date then;

$$\text{Study Day} = (\text{Assessment Date} - \text{Treatment Start Date}) + 1$$

All visit-based summaries will use nominal eCRF visits. Unscheduled visits will not be summarised in visit-based summaries.

Continuous data and ordered categorical data (if appropriate) will be summarised in terms of the number of non-missing observations, mean, standard deviation (SD), median, first and third quartiles, minimum and maximum and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, first and third quartiles will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarised in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.

Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator.

Changes from baseline in categorical data will be summarised using shift tables where appropriate.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. P-values less than 0.001 will be presented as "<0.001".

Confidence intervals will be presented to one more decimal place than the raw data.

All raw data will be listed.

All report outputs will be produced using SAS® version 9.3 or a later version in a secure and validated environment. All report outputs will be provided to NeRRe as Microsoft Word 2010 documents.

4.3 Study Subjects

Re-screening of subjects that screen fail during their first screening visit is allowed on this study. The decision regarding whether a particular subject is acceptable to re-screen

is made at the discretion of the PAREXEL Medical Monitor(s) and NeRRe. Subjects will be allowed to re-screen once. The decision for further re-screening will be determined on a case-by-case basis by NeRRe.

For subjects that re-screen, the screening data from their latest re-screening visit will be used for summary and analysis purposes. Only screening data from their latest re-screening visit will be listed.

During the analysis population classification meeting, a listing of screen failure reasons for subjects that were re-screened will be provided for review. A decision will be made during this meeting whether any re-screened subjects and/or re-screened subject data will be excluded from certain analyses. Any decisions made will be documented and approved by NeRRe.

4.3.1 Disposition of Subjects

A clear accounting of the disposition of all screened subjects will be provided, from screening to study completion.

The following subject disposition summaries will be provided:

- A summary of the number of subjects screened for entry into the study and the number and percentage of subjects excluded prior to randomisation by major reason and overall (Analysis population: All Subjects)
- A summary of the number of subjects randomised per centre (i.e. study site) and per region (Analysis population: All Subjects Randomised)
- A summary of the number of subjects randomised, the number and percentage of subjects treated (with at least one dose of study medication) and the number and percentage of subjects entering, withdrawing from study treatment, withdrawing from the study and attending each visit of the study by treatment group and overall. (Analysis population: All Subjects Randomised). Withdrawals from the study and from study treatment will also be summarised by major reason.

By-subject listings of eligibility details, randomisation details (including whether the blind was broken at discontinuation), visit dates and withdrawal/study completion details (including reason for discontinuation and duration of treatment prior to discontinuation) will be provided.

4.3.2 Protocol Deviations

Relevant protocol deviations are defined as those deviations from the protocol likely to have an impact on the perceived efficacy and/or safety of study treatments. The impact of relevant protocol deviations on the efficacy and/or safety results will be investigated by assessing the robustness of the study results and conclusions to the choice of analysis

population (see Section 4.4), both including and excluding data potentially affected by relevant protocol deviations.

Relevant protocol deviations and any action to be taken regarding the exclusion of subjects or affected data from specific analyses are defined in the project-specific Protocol Deviation Specification

A summary of the number and percentage of subjects with a relevant protocol deviation by treatment group and overall and by type of deviation (Analysis population: All Subjects Randomised) will be provided.

A by-subject listing of all protocol deviations will also be provided with a flag for relevant protocol deviations.

4.4 Analysis Populations

The efficacy summaries and analyses will be based on the **Full Analysis Set (FAS),** which is based upon the Intention-to-Treat (ITT) principle. The Full Analysis Set is defined as all randomised subjects who received at least one dose of double-blind study drug, irrespective of treatment received, who satisfy inclusion criterion number 4 (awake cough frequency of >=10 coughs/hour during the Screening period), and have awake cough frequency data from the ACM for at least one post-baseline assessment. In the event that a subject is allocated the incorrect study treatment as per the study randomisation list, subjects will be summarised and analysed 'as randomised' i.e. by randomised treatment group. In the event that a subject is stratified incorrectly, 'randomised stratum' will be used rather than 'actual stratum'.

For the primary efficacy variable(s), a sensitivity analysis will be performed on the **FAS** including subjects who fail to satisfy inclusion criterion number 4 (awake cough frequency of >=10 coughs/hour during the Screening period), to assess the robustness of the study conclusions to the choice of excluding subjects who failed to satisfy this inclusion criteria from the FAS.

For the primary efficacy variable(s), a sensitivity analysis will be performed on the **Per Protocol (PP) Set** to assess the robustness of the study conclusions to the choice of analysis population. The Per Protocol Set is defined as all subjects in the FAS excluding those identified as relevant protocol violators.

The safety summaries and analyses will be based on the **Safety Set**. The Safety Set is defined as all subjects who received at least one dose of double-blind study drug irrespective of treatment received. In the event that a subject is allocated the incorrect study treatment as per the study randomisation list, subjects will be summarised and analysed by the lowest dose of active study drug received following a worst-case approach.

The pharmacokinetic summaries and analyses will be based on the **Exposure Set**. The Exposure Set is defined as all subjects who received at least one dose of double-blind study drug and for whom at least one PK concentration is available. Subjects will be summarised by actual dose received. If Exposure-Response (ER) analysis is pursued and data permit, the exposure-response set will be defined in the Pop PK DAP and will be reported separately to the remainder of the study analyses.

Upon database release, protocol deviation and analysis population outputs will be produced and will be sent to NeRRe for review. An analysis population classification meeting will be arranged to discuss the outputs and to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects from analyses will be made prior to unblinding and will be documented and approved by NeRRe.

A summary of the number and percentage of subjects attending each study visit by treatment group and overall for each analysis population (Analysis population: All Subjects Randomised) will be provided.

A by-subject listing of analysis population details will also be provided. This listing will be presented by treatment group and will include: centre, subject identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All subjects screened will appear on this listing.

4.5 Demographic and Other Baseline Characteristics

Relevant Screening and Baseline data (i.e. data collected prior to the treatment administration) and demographic characteristics will be summarised descriptively for each treatment group. There will be no formal comparison of Baseline data, that is, no statistical hypothesis testing.

The following summaries of demographics and baseline characteristics will be provided:

- A summary of demographic variables (age, sex, race, ethnicity) by treatment group and overall.
- A summary of baseline disease factors (time since diagnosis of CRC [years], baseline awake objective cough frequency, baseline 24-hour objective cough frequency, baseline night-time objective cough frequency, baseline Cough Severity VAS, baseline Urge-to-Cough VAS, baseline LCQ score) by treatment group and overall.
- A summary of past and concomitant diseases and/or past surgeries by treatment group and overall.
- A summary of prior and concomitant medication by treatment group and overall.

Age will be calculated as the number of complete years between a subject's birth date and the date of informed consent.

Time since diagnosis of CRC will be derived as follows:

$$Time\ since\ diagnosis\ of\ CRC\ (Years) = \frac{(date\ of\ informed\ consent - date\ of\ diagnosis) + 1}{365.25}$$

If date of diagnosis is partially missing it will be imputed. If day is missing, this will be set to the first day of the month, "01". If month is missing, this will be set to the first month of the year, "January". If year is missing or the whole date is missing, then date of diagnosis will be left as missing.

Prior and concomitant medications will be coded using the WHO Drug Dictionary Version March 2017 or later.

Past and concomitant diseases and/or past surgeries will be coded using the MedDRA coding dictionary Version 20.0 or later.

Medication start and stop dates/times will be compared to the date/time of first dose of study medication to allow medications to be classified as either "Prior" or "Concomitant". Medications starting after the completion/withdrawal date will be listed but will not be classified or summarised.

Medications that start and stop prior to the date/time of first dose of study medication will be classified as "Prior".

Medications will be classified as "Concomitant" if they have a start date/time on or after the date/time of first dose of study medication or if they have a start date/time before the date /time of first dose of study medication and stop on or after the date/time of first dose of study medication (or are ongoing at study withdrawal).

If medication start and/or stop dates/times are missing or partial, the dates/times will be compared as far as possible with the date/time of first dose of study medication. Medications will be assumed to be "Concomitant", unless there is clear evidence (through comparison of partial dates/times) to suggest that the medication stopped prior to the first dose of study medication.

By-subject listings of demographic data and other baseline characteristics (as summarised above) will also be provided.

4.6 Treatment Compliance

The following measures will be used to calculate treatment compliance; the number of tablets dispensed at each visit and the number of tablets returned at each visit. The number of tablets taken at each visit will be calculated as the number of tablets dispensed

at the previous visit minus the number of tablets returned at the current visit. Treatment compliance will be calculated as follows:

$$Compliance\ (\%) = \left(\frac{Actual\ number\ of\ tablets\ taken}{Expected\ number\ of\ tablets\ taken}\right) x\ 100$$

The number of tablets expected to be taken is one tablet per day for each day in the study.

Non-compliant subjects will be defined as subjects with <80% compliance with study treatment administration. Compliant subjects will be defined as subjects with ≥80% compliance with study treatment administration.

A summary of the treatment compliance measures by treatment group and time interval, including the number and percentage of compliant and non-compliant subjects as per the definition above will be provided.

For the above summary, two time intervals will be defined:

1) Scheduled Duration: Compliance relative to the planned time on study i.e. the planned time on study was 12 weeks (85 days)
2) Actual Duration: Compliance relative to the actual time on study e.g. in the period between start of treatment date and end of treatment date.

A by-subject listing of treatment compliance data will also be provided.

4.7 Efficacy Evaluation

4.7.1 Analysis and Data Conventions

This study is designed to test for superiority. The null hypothesis for the treatment comparison will be that there is no difference between orvepitant treatment group and placebo in mean change in log transformed (to base 10) awake objective cough frequency at Week 12 compared to Baseline. The alternative hypothesis will be that there is a difference. Symbolically, this is expressed as follows:

H(0): μ(active) = μ(placebo)
H(1): μ(active) ≠ μ(placebo)

Two-sided tests with α=0.05 will be used to test this hypothesis. The tests will vary depending on the endpoint.

4.7.1.1 Multi-center Studies

For the purpose of the summaries and analyses, the term 'Centre' will be used to define each investigator site.

Due to region being used as a stratification factor for randomisation, the statistical model is already planned to adjust for differences between centres, by including 'Region' as a main effect term in the model, which will pool centres within the USA and within the UK for the North American and European regions, respectively.

4.7.1.2 Adjustments for Covariates

The primary efficacy analysis will be adjusted for the following baseline covariates:

- Region (stratification variable)
- Baseline log transformed awake objective cough frequency

An 'unadjusted' sensitivity analysis will be performed to assess the robustness of the study conclusions to the choice of baseline covariates.

In addition, for each covariate, a statistical test for the presence of a treatment-by-covariate interaction will be performed, by including the interaction term in the primary analysis model. If any of the treatment-by-covariate interactions are found to be statistically significant at the 10% level ($p<0.10$), subgroup analyses will be performed to further explore the nature of the interaction. In such cases, conclusions based on the primary analysis (no interaction) will be interpreted with caution.

4.7.1.3 Handling of Dropouts or Missing Data

Summary statistics will be based on non-missing values. For hypothesis tests, estimates and confidence intervals, missing values for continuous efficacy endpoints analysed via likelihood methods (e.g. repeated measures mixed models) will not be directly imputed as they are handled within the analysis itself, under the assumption that the model specification is correct and that the data is missing at random. The need for missing data imputation will be discussed during the first dry run and if required, will be detailed in a SAP Amendment prior to unblinding.

The number and percentage of subjects with missing data for the primary endpoint will be summarised by treatment group and overall.

A subject data listing will be provided showing all subjects with missing values for the primary endpoint. For these subjects, the listing will provide all observed data relating to the primary endpoint i.e. all measurements recorded prior to the missing value, any measurements recorded after the missing value, important baseline characteristics, the recorded reason(s) for study discontinuation and the timing of study discontinuation.

4.7.1.4 Multiple Comparisons/Multiplicity

One primary variable has been defined for this study (change from Baseline in awake objective cough frequency) and one time point of primary interest (Week 12). Multiple treatment contrasts are of interest (three in total: 10 mg versus placebo, 20 mg versus placebo and 30 mg versus placebo). Each comparison will be carried out at the two-sided 5% level of statistical significance. No adjustment will be made for the multiple orvepitant versus placebo group comparisons in this phase 2 study with the understanding that this increases the overall type I error rate of the study.

4.7.1.5 Interim Analyses

A sample size re-estimate was performed after at least 50 subjects had completed the Week 4 ACM assessment. The cough frequency data was reviewed to check the estimate of variability assumed for the primary efficacy variable. The interim sample-size re-estimation was done on a fully blinded basis using an overall pooled estimate of standard deviation. A revised standard deviation of change from Baseline (after taking logs) of 0.371 was calculated. This was an increase from the original assumption of 0.261.

It was noted during the review that the variance was disproportionately increased by four subjects who were retrospectively determined to be protocol violators of inclusion criteria 4 (IC04) as a result of re-analysis by Vitalograph of their baseline cough counts. A revised estimate of standard deviation of 0.285 was calculated after excluding these subjects. This was still an increase from the original assumption; however the effect was smaller than when the four IC04 violators were included.

Consequently, it was determined that a second sample size re-estimate would be performed after at least 100 subjects have completed the Week 12 ACM assessment. The sample size re-estimate will be performed separately for the FAS (see section 4.4) and for the FAS including subjects who fail IC04, so that the effect of these violators on a larger sample can be determined. The cough frequency data will be reviewed again to check the re-estimate of variability assumed for the primary efficacy variable. This interim sample-size re-estimation will be done on a fully blinded basis using an overall pooled estimate of standard deviation. Given that interim sample size re-estimation is limited to just an assessment of pooled variability and it is done on a fully blinded basis, the impact on type I error rate is expected to be negligible. See Section 4.10 for further details.

4.7.1.6 Examination of Subgroups

The uniformity of the treatment effect for the primary efficacy variable will be examined for the following subgroups:

- Region (North America, Europe)
- Sex (Male, Female)

The homogeneity of the treatment effect across subgroups will be investigated using graphical and/or analytical methods. A plot showing the mean treatment effect and 95%

confidence interval within each subgroup and overall will be provided. In addition, a statistical test for the presence of a treatment-by-subgroup interaction will be performed, by including the interaction term in the primary analysis model. If the treatment-by-subgroup interaction is found to be statistically significant at the 10% level ($p<0.10$), this will be taken as evidence of heterogeneity of the treatment effect across subgroups, and conclusions based on the primary analysis (no interaction) will be interpreted with caution.

Summaries of the primary efficacy variable by treatment group and subgroup will be produced. No formal statistical analysis will be performed within subgroup.

4.7.2 Primary Efficacy Variable – Change from Baseline to Week 12 in Awake Objective Cough Frequency

All primary efficacy summaries and analyses will be based upon the FAS as defined in Section 4.4. The analysis will be repeated using the PP set as a sensitivity analysis.

For the primary analysis of the primary endpoint at Week 12, there must be a minimum of 120 minutes (2 hours) of data (adjusted awake duration) at the week 12 visit for the assessments to be included using the FAS. For the primary analysis of the primary endpoint at Week 12, there must be a minimum of 240 minutes (4 hours) of data (adjusted awake duration) at the week 12 visit for the assessments to be included using the PP Set. All subjects, regardless of analysis population, will have at least 4 hours of awake data (adjusted awake duration) at screening (baseline) due to restrictions applied during the eligibility assessment performed by Vitalograph.

A separate sensitivity analysis using all randomised subjects (who received at least one dose of double-blind study drug and have awake cough frequency data from the ACM for at least one post-baseline assessment) will also be undertaken for the primary endpoint to assess the robustness of the results for the FAS where subjects who fail inclusion criteria 4 have been excluded.

Separate sensitivity analyses using all available post-baseline data (regardless of minimum duration) will also be undertaken for all cough frequency endpoints.

The primary variable for the assessment of efficacy is the change from Baseline to Week 12 in awake objective cough frequency measured with an ACM, the 'Vitalojak™'. The awake duration is considered to be the time between session start and end time, excluding time when subject was flagged as asleep. This can be determined from the 24-hour acoustic recordings and approximates to waking hours when cough rates are highest.

To evaluate awake cough frequency, the following definition will be used:

$$Awake\ Cough\ Frequency\ (coughs\ /\ hr) = (\frac{Adjusted\ Awake\ Cough\ Count}{Adjusted\ Awake\ Duration(s)})\ x\ 3600$$

See Section 6.1 for definitions of adjusted cough count and adjusted duration.

Awake objective cough frequency will be summarised by treatment group, and visit (Baseline, Weeks 2, 4 and 12) in terms of absolute values and changes from baseline on the original scale (including percent change from baseline) and after log (base 10) transformation. If there are any zero frequencies, a nominal amount (0.1) will be added to all frequencies prior to log transformation.

The effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2, Week 4 and Week 12) in log transformed awake objective cough frequency will be analysed using a mixed model for repeated measures (MMRM). The MMRM model will include region, baseline log-transformed awake objective cough frequency, treatment, week, week by region interaction, week by baseline interaction and week by treatment interaction as explanatory variables. Subject will be included as a random effect. Restricted maximum likelihood (REML) estimation will be used.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. If there are still issues with the fit of the model or estimation of the treatment effects, subject will be treated as a fixed effect.

The treatment effect at Week 12 will be estimated using the ratio of geometric means (i.e. the difference between the treatments least squares means [adjusted means] on the log scale, back-transformed to the original scale). The adjusted geometric mean (GM) ratio (Week 12/ Baseline), ratio of GM ratios (Active/ Placebo), 95% confidence interval for the ratio of GM ratios and the p-value from the hypothesis test of no difference between the treatment groups at Week 12 will be presented. The ratio of GM ratios is the geometric mean ratio (Active/ Placebo) of the Week 12/ Baseline result.

A plot showing the adjusted GM ratio (Week X/ Baseline) in awake objective cough frequency over time within each treatment group will be provided.

A scatter plot showing the change from baseline to week 12 in awake objective cough frequency versus baseline awake cough frequency for individual subjects within each treatment group will also be provided.

Diagnostic plots will be presented to assess the suitability of the model.

The sensitivity analyses using all available post-baseline data (regardless of minimum duration) and all randomised subjects will also be undertaken for change from Baseline in log transformed awake objective cough frequency using the MMRM model specified previously.

In the event of zero counts, as a sensitivity analysis, the effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2, Week 4 and Week 12) in non-transformed adjusted awake objective cough count will be analysed using a generalised linear mixed model (GLMM) for repeated measures using a negative binomial distribution with the adjusted awake duration (hours) as an offset variable. The GLMM model will include region, baseline awake objective cough count, treatment, week, week by region interaction, week by baseline interaction and week by treatment interaction as explanatory variables. Subject will be included as a random effect. An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. If there are still issues with the fit of the model or estimation of the treatment effects, subject will be treated as a fixed effect.

The robustness of the study results to the choice of baseline covariates will be investigated as described in Section 4.7.1.2. In addition, the presence of treatment-by-covariate interactions will be investigated.

Further sensitivity analyses to investigate the impact of the choice of method for handling missing data and the choice of analysis population will be performed as described in Section 4.7.1.3.

The homogeneity of the treatment effect for a number of important subgroups (i.e. region, sex) will be investigated as described in Section 4.7.1.6

A by-subject listing of the primary efficacy data will be provided.

Dose-response will be evaluated in the context of the primary MMRM model for change from baseline in awake objective cough frequency. A plot showing the adjusted GM ratio (Week X/ Baseline) and 95% CI in awake objective cough frequency at each dose by visit will be provided. A scatter plot showing the change from baseline in awake objective cough frequency at each dose by visit will be provided. The presence of a linear dose response relationship will be investigated. The presence of a non-linear dose response will be investigated if suggested by the plot of the data.

4.7.3 Secondary Efficacy Variables

All secondary efficacy summaries and analyses will be based upon the FAS as defined in Section 4.4.

In the analyses of the secondary efficacy endpoints that use cough frequency, there must be a minimum of 120 minutes (2 hours) of data (adjusted endpoint duration) for that endpoint at the relevant post-baseline visit for the assessment to be included. All subjects, regardless of analysis population, will have at least 4 hours of awake data (adjusted awake duration) at screening (baseline) due to restrictions applied during the eligibility assessment performed by Vitalograph.

Change in awake objective cough frequency at Weeks 2, and 4 compared to Baseline

The analysis of change from Baseline in awake objective cough frequency at Weeks 2 and 4 will be conducted as part of the analysis for the primary endpoint described in Section 4.7.2. Change from baseline summaries and the results from the MMRM analysis (adjusted GM ratio (Week X/ Baseline), ratio of GM ratios (Active/ Placebo), 95% confidence interval for the ratio of GM ratios and the p-value from the treatment comparison at Weeks 2 and 4) will be presented.

A sensitivity analysis using all available post-baseline data (regardless of minimum duration) will also be undertaken for change from Baseline in log transformed awake objective cough frequency using the MMRM model specified previously.

In the event of zero counts, as a sensitivity analysis, the effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2 and Week 4) in non-transformed adjusted awake objective cough count will be analysed using a GLMM for repeated measures using a negative binomial distribution with the adjusted awake duration (hours) as an offset variable.

Change in 24-hour objective cough frequency at Weeks 2, 4 and 12 compared to Baseline

To evaluate 24-hour cough frequency, the following definition of total cough frequency (coughs/hr) will be used:

$$Total\ Cough\ Frequency\ (coughs\ /\ hr) = \left(\frac{Adjusted\ Total\ Cough\ Count}{Adjusted\ Total\ Recording\ Duration(s)}\right) x\ 3600$$

See Section 6.1 for definitions of adjusted cough count and adjusted duration.

The analysis of change from Baseline in 24-hour objective cough frequency will be analysed in the same way as for the primary endpoint as described in Section 4.7.2. Change from baseline summaries and the results from the MMRM analysis (adjusted GM ratio (Week X/ Baseline), ratio of GM ratios (Active/ Placebo), 95% confidence interval

for the ratio of GM ratios and the p-value from the treatment comparison at Weeks 2, 4 and 12) will be presented.

A sensitivity analysis using all available post-baseline data (regardless of minimum duration) will also be undertaken for change from Baseline in log transformed 24-hour objective cough frequency using the MMRM model specified previously.

In the event of zero counts, as a sensitivity analysis, the effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2, Week 4 and Week 12) in non-transformed adjusted total objective cough count will be analysed using a GLMM for repeated measures using a negative binomial distribution with the adjusted total recording duration (hours) as an offset variable.

Change in night-time (non-waking time) objective cough frequency at Weeks 2, 4 and 12 compared to Baseline

To evaluate night time cough frequency, the following definition of asleep cough frequency (coughs/hr) will be used:

$$Asleep\ Cough\ Frequency\ (coughs\ /\ hr) = \left(\frac{Adjusted\ Asleep\ Cough\ Count}{Adjusted\ Asleep\ Recording\ Duration(s)}\right) x\ 3600$$

See Section 6.1 for definitions of adjusted cough count and adjusted duration.

The analysis of change from Baseline in night-time (non-waking time) objective cough frequency will be analysed in the same way as for the primary endpoint as described in Section 4.7.2. Change from baseline summaries and the results from the MMRM analysis (adjusted GM ratio (Week X/ Baseline), ratio of GM ratios (Active/ Placebo), 95% confidence interval for the ratio of GM ratios and the p-value from the treatment comparison at Weeks 2, 4 and 12) will be presented.

A sensitivity analysis using all available post-baseline data (regardless of minimum duration) will also be undertaken for change from Baseline in log transformed night-time objective cough frequency using the MMRM model specified previously.

In the event of zero counts, as a sensitivity analysis, the effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2, Week 4 and Week 12) in non-transformed adjusted asleep objective cough count will be analysed using a GLMM for repeated measures using a negative binomial distribution with the adjusted asleep recording duration (hours) as an offset variable.

Change in the Cough Severity visual analogue scale (VAS) at Weeks 2, 4, 8, and 12 compared to Baseline

The cough severity VAS is a 100-mm scale on which subjects indicate the severity of cough in the previous 24 hours, both during the daytime (awake time) and during night-time (non-awake time) separately. The VAS ranges from "no cough" on the left to "worst cough" on the right.

Change in the Cough Severity VAS will be summarised by treatment group and visit (Weeks 2, 4, 8 and 12) in terms of absolute values and changes from baseline.

The effect of treatment in terms of the change from Baseline to each scheduled visit (Week 2, Week 4, Week 8 and Week 12) in Cough Severity VAS will be analysed using a MMRM. The MMRM model will include region, baseline cough severity VAS, treatment, week, week by region interaction, week by baseline interaction and week by treatment interaction as explanatory variables. Subject will be included as a random effect. Restricted maximum likelihood (REML) estimation will be used.

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, and autoregressive. If there are still issues with the fit of the model or estimation of the treatment effects, subject will be treated as a fixed effect.

The treatment effect at Week 12 will be estimated using the difference between the treatments least squares means (adjusted means). The 95% confidence interval for the difference between the treatments least squares means and the p-value from the hypothesis test of no difference between the treatment groups will be presented.

A plot showing the adjusted mean change from baseline in Cough Severity VAS over time within each treatment group will be provided.

Diagnostic plots will be presented to assess the suitability of the model.

<u>Change in the Urge-to-Cough VAS at Weeks 2, 4, 8, and 12 compared to Baseline</u>
The Urge-to-Cough VAS is a 100-mm scale on which subjects indicate their Urge-to-Cough over the previous 24 hours (day/awake time and night time combined). The VAS ranges from "no Urge-to-Cough" on the left to "severe Urge-to-Cough" on the right.

The analysis of change from Baseline in Urge-to-Cough VAS will be analysed in the same way as for the secondary endpoint, Cough Severity VAS, as described above.

<u>Change in the Leicester Cough Questionnaire (LCQ) score (total and three domain scores) at Weeks 2, 4, 8, and 12 compared to Baseline</u>
The LCQ is a 19-item questionnaire that assesses cough-related Quality-of-Life. It has three domains (physical, psychological and social). The total score range is 3-21 and

domain scores each range from 1-7; a higher score indicates a better Quality-of-Life. For incomplete questionnaires, the 10% criteria for missing data will be applied. See Section 6.2 for further details on the scoring of the LCQ and handling of missing data.

The analysis of change from Baseline in LCQ total score and the three domain scores will be analysed in the same way as for the secondary endpoint, Cough Severity VAS, as described above.

Global Rating of Change in Cough Frequency and Severity at Weeks 2, 4, 8, and 12
In the Global Rating of Change Scale, subjects will indicate if there has been a change in their symptoms (cough frequency and, separately, cough severity) since starting the Investigational Medicinal Product (IMP). They will respond with "worse", "about the same" or "better".

These will be summarised by treatment group and visit (Weeks 2, 4, 8 and 12) as categorical endpoints ("worse", "about the same" and "better"). Pairwise comparisons will be performed using the Cochran-Mantel-Haenszel test stratified by region using modified ridit scores (row mean scores differ).

If subjects indicate a change (either "worse" or "better") they will then further indicate on a 7-point scale the degree of change ranging from 1 (almost the same, hardly any change) to 7 (a very great deal changed). This will also be summarised as a categorical endpoint by change category (worse, better), treatment group and visit.

4.8 Safety Evaluation

All safety summaries and analyses will be based upon the Safety Set as defined in Section 4.4.

Safety data will be summarised descriptively for each treatment group. No formal inferential tests will be performed on safety data.

4.8.1 Extent of Exposure

The extent of exposure (days) will be derived as follows:

$$Duration\ of\ exposure\ (days) = Date\ of\ last\ dose - Date\ of\ first\ dose + 1$$

The following extent of exposure summaries will be provided:

- A summary of the duration of exposure to treatment, by treatment group. The mean, median, and the number and percentage of subjects exposed during the 12-week treatment period will be provided.

A by-subject listing of exposure data will also be provided.

4.8.2 Adverse Events

Adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 20.0 or higher.

Treatment-emergent adverse events (TEAEs) will be tabulated and are defined as those adverse events that either first occur or worsen in severity on or after the date/time of first dose of IMP.

Where dates or times are missing or partially missing, adverse events will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the adverse event started prior to the first dose of study treatment.

Summaries will present both the number and percentage of subjects reporting an AE and the number of AEs reported.

The following adverse event summaries will be provided:

- An overall summary of adverse events by treatment group
- A summary of the number and percentage of subjects reporting an adverse event by treatment group, system organ class (SOC), and preferred term (PT)
- A summary of the number and percentage of subjects reporting a TEAE by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting a treatment-related TEAE by treatment group, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE by treatment group, severity, SOC, and PT

Adverse event summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the orvepitant treatment group, and then similarly by decreasing frequency in the placebo treatment group, and then alphabetically for SOC, and PT within SOC.

For each subject and each adverse event, the worst severity recorded will be attributed and used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, the worst case will be assumed.

A by-subject listing of all adverse events (including non-treatment-emergent events) will be provided. This listing will be presented by treatment group and will include: centre, subject identifier, age, sex, race, date/time of first dose of IMP, adverse event (SOC, PT, and verbatim term), treatment-emergent flag, date/time of onset, date/time of resolution, duration, severity, seriousness, action taken, outcome and causality.

4.8.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events

The following adverse event summaries will be provided:

- A summary of the number and percentage of subjects with fatal TEAEs during the study, by treatment group, SOC and PT
- A summary of the number and percentage of subjects reporting a serious TEAE, by treatment group, SOC and PT
- A summary of the number and percentage of subjects with TEAEs leading to discontinuation of study treatment, by treatment group, SOC and PT

The following listings will also be provided:

1) A by-subject listing of all fatal adverse events during the study
2) A by-subject listing of all serious adverse events
3) A by-subject listing of all adverse events leading to discontinuation of study treatment

Listings will follow the format described for adverse events in Section 4.8.2.

4.8.4 Clinical Laboratory Evaluation

Blood for clinical chemistry assessments will be collected as indicated in the Study Schedule in **Table 1: Schedule of Events**. The following clinical chemistry parameters will be assessed: sodium, potassium, glucose, urea, creatinine, creatine kinase, albumin, calcium, phosphate, bilirubin (total), alkaline phosphatase, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma glutamyl transferase (GGT), bicarbonate, magnesium, chloride, pregnancy (serum), creatinine clearance (est.) and total protein.

Blood for hematology assessments will be collected as indicated in in the Study Schedule in **Table 1: Schedule of Events**. The following hematology parameters will be assessed: red blood cell (RBC) count, white blood cell (WBC) count, hematocrit, hemoglobin, MCV, platelet count and WBC differentials (neutrophils, total lymphocytes, monocytes, eosinophils and basophils) [absolute and %].

Urine for urinalysis assessments will be collected as indicated in the Study Schedule in **Table 1: Schedule of Events**. Urinalysis will include glucose, bilirubin, ketones, specific gravity, hemoglobin, erythrocytes, leucocytes, pH, protein, urobilinogen, nitrites, leucocyte esterase, hyaline casts, granular casts, waxy casts, WBC casts, RBC casts, epithelial cells, crystals, mucous threads, bacteria and yeast. A urine sample will also be collected to perform the opiate drug screen.

For by-visit summaries, the last non-missing assessment (including repeat assessments) recorded at each visit will be summarised.

The following summaries will be provided:

- A summary of each continuous laboratory parameter by treatment group and visit.

- A summary of the change from baseline in each continuous laboratory parameter by treatment group and visit.
- A summary of the number and percentage of subjects experiencing low, normal or high values at baseline and at selected post-baseline time points, by laboratory parameter and treatment group (shift table)
- A summary of opiate screen results (positive/ negative) by treatment group and visit.
- A summary of categorical laboratory parameter at baseline and at selected post-baseline time points, by laboratory parameter and treatment group (shift table)

A by-subject listing of all laboratory data will be provided by treatment group, with abnormal values highlighted, and including centre, subject identifier and visit. Laboratory reference ranges will also be listed.

4.8.5 Vital Signs, Physical Findings and Other Observations Related to Safety

Vital signs will be measured at the time points indicated in **Table 1: Schedule of Events** and will include systolic and diastolic BP, pulse, temperature and weight. Weight will be measured at Baseline/Day 1, Week 12 and Week 14 Follow-up only.

12-lead ECG variables will be measured at the time points indicated in **Table 1: Schedule of Events** and will include heart rate and rhythm and RR, PR, QRS, QT, QTcF and QTcB intervals. 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.

The following summaries will be provided:

- A summary of each vital sign parameter by treatment group and visit
- A summary of the change from baseline in each vital sign parameter by treatment group and visit.
- A summary of each continuous ECG parameter (*heart rate, RR, PR, QRS, QT, QTcF and QTcB)* by treatment group and visit
- A summary of the change from baseline in each continuous ECG parameter (*heart rate, RR, PR, QRS, QT, QTcF and QTcB)* by treatment group and visit.
- A summary of QTcF and QTcB by category (<450, 450 to <480, 480 to <500, ≥500 msec) and treatment group
- A summary of the change from baseline in QTcF and QTcB by category (<0, 0 to <30, 30 to <60, ≥60 msec) and treatment group.
- A summary of the number and percentage of subjects experiencing normal, abnormal – not clinically significant or abnormal – clinically significant ECG by treatment group and visit
- A summary of the number and percentage of subjects experiencing normal, abnormal – not clinically significant or abnormal – clinically significant ECG at baseline and at selected post-baseline time points, by treatment group (shift table)

By-subject listings of vital sign parameters, ECG results, physical examination results and any other observations related to safety will also be provided. Heart rhythm will be listed only.

4.9 Other Analyses

4.9.1 Pharmacokinetics

All PK summaries and analyses will be based upon the orvepitant concentration and time data as defined in Section 4.4 and if data permit, an Exposure-Response Set as defined in Section 4.4.

If data permit and POP PK is pursued, a PK exposure-response relationship for the orvepitant groups (10 mg, 20 mg, and 30 mg) will be carried out to examine the possible relationship over time between the primary endpoint of clinical efficacy and plasma levels of the drug.

The plasma orvepitant concentrations will be summarised by treatment group and visit (Weeks 2, 4, 8 and 12). The following summary statistics will be presented; number of non-missing observations, mean, SD, coefficient of variation (CV%), median, minimum, maximum, geometric mean, CV% for geometric mean, and number of subjects below the lower limit of quantification (LLoQ). LLoQ will be substituted for below lower limit of quantification (BLLQ) values.

If appropriate, the plasma orvepitant concentrations will be analysed using a population pharmacokinetic (Pop PK) approach by nonlinear mixed effects modelling (NONMEM®). The Pop PK model will be used to evaluate the relationship between orvepitant exposure and clinical endpoints. The details of the analysis will be described in a separate Pop PK Data Analysis Plan (DAP).

4.10 Determination of Sample Size

The initial planned sample size for this study was 55 subjects per treatment group (220 subjects in total). The sample size calculation was performed in SAS® 9.4, based on pairwise comparisons of the primary endpoint for each active dose versus placebo. Objective cough frequency will be analysed on a log scale (base 10). A two-sided type I error of 0.05, not adjusted for multiple comparisons against placebo, will be used for this exploratory study. It was assumed a standard deviation of change from Baseline (after taking logs) of 0.261 (based on previous study data), a reduction in the placebo group of 10% and a reduction in the orvepitant group of 35%, 55 subjects per arm would provide a power of 80%.

The SAS code for the sample size calculation is as follows:

proc power;

twosamplemeans test=diff
meandiff = **-0.141**
stddev = **0.261**
sides = **2**
alpha = **0.05**
ntotal = **.**
power = **0.8**;
run;

As a result of the planned sample size re-estimation the sample size was increased to 65 subjects per group (260 in total). Assuming a revised standard deviation of change from Baseline (after taking logs) of 0.285 (based on the sample size re-estimate and increased from the original assumption of 0.261) a reduction in the placebo group of 10% and a reduction in the orvepitant group of 35%, 65 subjects per arm will provide a power of 80%. This revised estimate of standard deviation was obtained (at the first sample size review) based on the Full Analysis Set (see Section 4.4).

After at least 100 subjects have completed the Week 12 ACM assessment (subsequent to the now completed first sample size re-estimate) the cough frequency data will be reviewed to check the estimate of variability assumed for the primary efficacy variable. This interim sample-size re-estimation will be done on a fully blinded basis using an overall pooled estimate of standard deviation, and may result in an increase in the sample size (Kieser and Friede, 2000). Only the awake cough frequency data will be reviewed.

The estimate of standard deviation will be obtained using all available data at the time of the interim. The following model will be used to derive the overall pooled estimate of standard deviation:

proc mixed data=d;
class subjid week;
model chg = week base region region*week base*week;
repeated week / sub=subjid type=vc;
run;

The impact of the choice covariance matrix will be investigated.

The sample size will then be re-estimated using East 6 software.

Assuming no further changes to the sample size estimate, approximately 292 subjects will be randomised to ensure at least 260 completers (assuming approximately 10% drop out rate). If the sample size re-estimation indicates that more subjects are required, then up to 320 subjects (approximately 80 per group) may be recruited; the precise number will be documented in a non-substantial (administrative) amendment to the protocol. Any possible impact on the type I error due to this blinded sample size re-assessment procedure is believed to be negligible (Kieser and Friede, 2003).

4.11 Changes in the Conduct of the Study or Planned Analysis

In section 11.2 of the Protocol, the Exposure-Response (ER) Set is defined as all subjects who received at least one dose of double-blind study drug and for whom the PK data are considered sufficient. For the purposes of the PK analysis scoped within this SAP, an Exposure Set is defined (see section 4.4 for definition) for a PK summary table. Exposure-response analyses, if performed, will be within scope of a separate Pop PK DAP and the analysis population will be defined therein accordingly.

In section 8.4 of the Protocol, respiration rate is stated as a safety endpoint but as per the schedule of events, this parameter is not to be collected as part of vital signs and therefore will not be included in the summary of vital signs.

5 REFERENCES

Journals:

[1] Kieser M, Friede T. Blinded sample size re-estimation in multi-armed clinical trials. Drug Information Journal 2000; 34:455– 460.

[2] Kieser M, Friede T. Simple procedures for blinded sample size adjustment that do not affect the type 1 error rate. Statist. Med 2003; 22:3571-3581

[3] Birring S, Carr A, Singh S, Morgan ML, Pavord I. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). Thorax 2003; 58:339-343

[4] Abdulqawi R, Dockry R, Holt K, McCarthy B, Ford A, Smith J. P2X3 receptor antagonist (AF-219) in refractory chronic cough: a randomised, double-blind, placebo-controlled phase 2 study. Lancet 2014, 14:61255-1

6 APPENDICES

6.1 Appendix 1 – Cough Frequency Definitions

Data Provided

- Start and end times of recording
- Events: each cough is a discrete event with an associated event/cough time
- Asleep start and end times flags
- Comments (and flag times) on recorder being removed or turned off (N.B. only periods when the device is obviously removed will be flagged, periods of "high volume" will be included in the durations and in the analysis)
- Comments on data quality

Definitions

Session	A protocol required ACM measurement period over 24 hours (screen, repeat screen, Week 2, Week 4, Week 12)
Session Start Time	Earliest accepted time for the session
Session End Time	End time of the recording for the session (this should be ≤24 hours after the Start Time)
Total Session Duration (*Seconds)*	= Session End Time – Session Start Time + 1

Sleep Start Time	Time flagged that subject is asleep.
Awake Start Time	Time flagged that subject is awake, having previously had a flag indicating asleep
Asleep Duration (*Seconds)*	Sum of durations for all asleep periods during the session. Duration of a single asleep period will be calculated as: Awake Start Time – Sleep Start Time + 1
Awake Duration (*Seconds)*	= Total Session Duration – Asleep Duration

Off Time	Time flagged when the recorder is unequivocally turned off during the session. If the recorder is turned off it will be assumed that it is not turned back on again and the "Off Time" becomes the "Session End Time".

Device Removed Time	Time flagged that recorder is removed from subject.
Device Restored Time	Time flagged that recorder is put back on, having previously had a flag indicating recorder has been removed.
Removed Duration (*Seconds)*	= Device Restored Time – Device Removed Start Time + 1. Time during which the recorder is unequivocally removed from the subject during the session, as flagged. A subject may have

	more than one Removed Period during a Session.

Total Adjustment Duration (*Seconds)*	Total of all "Removed Durations" for Removed Periods that occurred during the "Total Session Duration", that is, the "Device Removed Time" occurred after a "Session Start Time" and the "Device Restored Time" occurred before a "Session End Time".
Asleep Adjustment Duration (*Seconds)*	Total of all "Removed Durations" for Removed Periods that occurred during any asleep period, that is, the "Device Removed Time" occurred after a "Sleep Start Time" and the "Device Restored Time" occurred before an "Awake Start Time".
Awake Adjustment Duration (*Seconds)*	Awake Adjustment Duration = Total Adjustment Duration - Asleep Adjustment Duration

Adjusted Total Recording Duration (*Seconds)*	= Total Session Duration – Total Adjustment Duration
Adjusted Awake Duration (*Seconds)*	= Awake Duration – Awake Adjustment Duration
Adjusted Asleep Duration (*Seconds)*	= Asleep Duration – Asleep Adjustment Duration

Total Cough Count	Number of events (coughs) recorded between the "Session Start Time" and "Session End Time"
Asleep Cough Count	Number of events (coughs) recorded during all asleep periods. That is the event occurred between a "Sleep Start Time" and an "Awake Start Time".
Awake Cough Count	= Total Cough Count – Asleep Cough Count

Total Adjustment Count	Total number of events (coughs) recorded during all "Device Removed Periods" that occurred during the "Total Adjustment Duration".
Asleep Adjustment Count	Total number of events (coughs) recorded for all for "Device Removed Periods" that occurred during the "Asleep Adjustment Duration".
Awake Adjustment Count	Awake Adjustment Count = Total Adjustment Count - Asleep Adjustment Count

Adjusted Total Cough Count	= Total Cough Count – Total Adjustment Count
Adjusted Awake Cough Count	= Awake Cough Count – Awake Adjustment Count
Adjusted Asleep Cough Count	= Asleep Cough Count – Asleep Adjustment Count

Eligibility Assessment

- To be eligible for randomisation a subject must have a mean awake cough frequency of ≥10 coughs/hr
- To determine eligibility:
  - Recording must be technically acceptable, and
  - There must be a minimum of 240 minutes (4 hours) of awake data (adjusted awake duration)

If a Session meets these criteria, then:

  - Cough frequency will be determined from the Screening (or re-screening) recording based on the Awake Cough Frequency (coughs/hr):

= (Adjusted Awake Cough Count / Adjusted Awake Duration (s)) x 3600

- Feedback to site will be one of the following 3 options:
  1. Awake Cough frequency ≥10/hour
  2. Awake Cough frequency <10/hour
  3. Recording rejected

Baseline Assessment

The screening recording used to determine eligibility will also be considered as the baseline assessment for analyses. Therefore, all randomised subjects included in analysis will have at least 4 hours of awake data (adjusted awake duration) due to the restrictions applied above.

Awake Cough Frequency Efficacy Assessments

The primary efficacy and some secondary efficacy assessments are based on awake cough frequency.

- The Awake Cough Frequency (coughs/hr) will be used for these analyses (as defined above)
- There must be a minimum of 120 minutes (2 hours) of awake data (adjusted awake duration) at the week 12 visit for the assessments to be included in the primary analysis of the primary endpoint at Week 12 using the Full Analysis Set
- There must be a minimum of 240 minutes (4 hours) of awake data (adjusted awake duration) at the week 12 visit for the assessments to be included in the analysis of the primary endpoint at Week 12 based using the Per Protocol Set.
- There must be a minimum of 120 minutes (2 hours) of awake data (adjusted awake duration) at the relevant post-baseline visit for the assessment to be included in the analyses of the secondary efficacy endpoints that use awake cough frequency
- Sensitivity analyses using all available post-baseline data (regardless of minimum duration) will also be undertaken for all awake cough frequency endpoints

24 Hour Cough Frequency Assessments

- Secondary efficacy assessments that are based on 24-hour cough frequency will use the Total Cough Frequency (coughs/hr):

= (Adjusted Total Cough Count / Adjusted Total Recording Duration (s)) x 3600

- There must be a minimum of 120 minutes (2 hours) of 24-hour data (adjusted total duration) at the relevant post-baseline visit for the assessment to be included in the analyses of the secondary efficacy endpoints that use 24-hour cough frequency
- Sensitivity analyses using all available post-baseline data (regardless of minimum duration) will also be undertaken for the 24-hour cough frequency endpoint

Night Time Cough Frequency Assessments

- Secondary efficacy assessments that are based on night time cough frequency will use the Asleep Cough Frequency (coughs/hr):

= (Adjusted Asleep Cough Count / Adjusted Asleep Recording Duration (s)) x 3600

- There must be a minimum of 120 minutes (2 hours) of asleep data (adjusted asleep duration) at the relevant post-baseline visit for the assessment to be included in the analyses of the secondary efficacy endpoints that use night time cough frequency
- Sensitivity analyses using all available post-baseline data (regardless of minimum duration) will also be undertaken for the night-time cough frequency endpoint

Other General Principles

- Average cough frequency will be reported to 2 decimal places
- An event/cough is allocated according to its time (recorded as hours:minutes:seconds) and the corresponding interval start and stop times.
    - 24-hour cough if:
        - Session Start Time <= Event Time < Session End Time
        - Cough is excluded from analysis if its time falls within a removed duration, i.e. if Device Removed Time <= Event Time < Device Restored Time
    - Awake cough if:
        - Session Start Time <= Event Time < Sleep Start time
        - Awake Start Time <= Event Time < Session End time
        - Cough is excluded from analysis if its time falls within a removed duration, i.e. if Device Removed Time <= Event Time < Device Restored Time
    - Asleep cough if:
        - Sleep Start Time <= Event Time < Awake Start time
        -
        - Cough is excluded if its time falls within a removed duration, i.e. if Device Removed Time <= Event Time < Device Restored Time

Reference: Provided by NeRRe

6.2 Appendix 2 – Scoring of LCQ

1) Domains (questions):
    a) Physical: 1,2,3,9,10,11,14,15
    b) Psychological: 4,5,6,12,13,16,17
    c) Social: 7,8,18,19

2) Domain scores: total score from items in domain/number of items in domain (range 1-7).

3) Total scores: addition of domain scores (range 3-21).

MISSING DATA
Missing data should be avoided where possible. The investigator can choose the most appropriate method to handle missing and this should be stated clearly in the study report. Below are 2 suggestions, the 10% criteria is our preference.

MISSING DATA (10% criteria)
Our preference for scoring when there is missing data:
Physical domain: only 1 missing item allowed. (2 missing items = cannot calculate score)
Psychological domain: only 1 missing item allowed.
Social domain: no missing items allowed.
Total score: always requires all 3 domain scores.
The average score for items available for a specific domain is used as the value for the missing item.

MISSING DATA (25% criteria)
Some investigators use a more liberal definition for missing items.
Physical domain: only 2 missing items allowed. (3 missing items = cannot calculate score)
Psychological domain: only 1 missing item allowed.
Social domain: only 1 missing item allowed.
Total score: always requires all 3 domain scores.
The average score for items available for a specific domain is used as the value for the missing item.

Reference: Birring et al, 2003 and private correspondence with S Birring (see attached email)

FW LCQ missing data.msg