STATISTICAL ANALYSIS PLAN

PROTOCOL TITLE:	A double-blind, randomised, placebo controlled, two period cross-over study to evaluate the efficacy and safety of orvepitant in chronic cough in patients with idiopathic pulmonary fibrosis
PROTOCOL (Short Name, Version, Date):	IPF COMFORT (ORV-PF-01), Versions 3.0/3.1, 05 April 2023
STUDY DRUG:	Orvepitant
STUDY PHASE:	Phase II
SPONSOR:	NeRRe Therapeutics Ltd
SAP DATE	14 March 2024
STATUS	Final v1.0

TABLE OF CONTENTS

LIST OF ABBREVIATIONS **6**

1 PURPOSE **8**

1.1 RESPONSIBLITIES 8

2 INTRODUCTION **8**

2.1 STUDY OBJECTIVES 9
2.1.1 Primary Objective(s) *9*
2.1.2 Secondary Objective(s) *9*
2.1.3 Exploratory Objective(s) *9*
2.2 STATISTICAL HYPOTHESIS 9
2.3 STUDY POPULATION 9
2.4 TRIAL DESCRIPTION 10
2.5 STUDY SAMPLE SIZE DETERMINATION 10
2.5.1 Estimated Sample Size *10*
2.5.2 Sample Size Re-Estimation *11*
2.6 TREATMENT ASSIGNMENT AND BLINDING 11
2.6.1 Randomisation *11*
2.6.2 Blinding *12*
2.6.3 Planned and Actual Treatment *12*
2.6.4 Carry Over Effects *12*
2.7 ADMINISTRATION OF TRIAL MEDICATION 12
2.7.1 Missed dose *13*
2.7.2 Dose interruption *13*
2.8 SCHEDULE OF ASSESSMENTS AND PROCEDURES 14

3 ENDPOINTS **16**

3.1 PRIMARY EFFICACY ENDPOINT AND ESTIMAND 16
3.1.1 ESTIMAND *16*
3.2 SECONDARY EFFICACY ENDPOINTS 17
3.3 EXPLORATORY ENDPOINTS 18
3.4 SAFETY ENDPOINTS 19

4 ANALYSIS POPULATIONS **20**

4.1 SAFETY SET (SS) 20
4.2 FULL ANALYSIS SET (FAS) 21
4.3 PER PROTOCOL SET (PPS) 21
4.5 PROTOCOL DEVIATIONS 21

5 GENERAL ASPECTS FOR STATISTICAL ANALYSES **21**

5.1 GENERAL METHODS 21
5.2 KEY DEFINITIONS 22
5.2.1 Study Day *22*
5.2.2 Baseline Values *22*
5.3 DATA HANDLING CONVENTIONS 23

5.3.1 Missing Data *23*
5.3.2 Visit Windowing *23*
5.3.3 Data Dates *25*
5.3.4 Pooling of Sites *25*

6 STUDY SAMPLE **25**

6.1 DISPOSITION OF SUBJECTS 25
6.2 DEMOGRAPHICS AND BASELINE CHARACTERISTICS 25
6.3 MEDICAL HISTORY 25
6.4 PRIOR AND CONCOMITANT MEDICATIONS 26

7 EFFICACY ANALYSES **27**

7.1 PRIMARY AND SECONDARY EFFICACY ANALYSIS 27
7.1.1 Primary Efficacy Endpoint *27*
7.1.2 Secondary Efficacy Endpoints *29*
7.1.3 Sub-Groups *33*

8 SAFETY ANALYSES **34**

8.1 ADVERSE EVENT 34
8.2 CLINICAL LABORATORY TESTS 36
8.3 VITAL SIGNS 36
8.4 ECG DATA 36
8.5 PHYSICAL EXAMINATION 37
8.6 PULMONARY FUNCTION TESTS 37
8.7 TREATMENT COMPLIANCE AND EXPOSURE 37

9 BIOMARKERS **39**

10 PHARMACOKINETIC ANALYSES **39**

11 INTERIM ANALYSIS **39**

11.1 DATA SAFETY MONITORING BOARD (DSMB) 40

12 SOFTWARE AND PROGRAMMING SPECIFICATIONS **40**

12.1 GENERAL PROGRAMMING SPECIFICATIONS 40
12.2 TABLE, LISTING, AND FIGURE FORMAT 40
12.2.1 General *40*
12.2.2 Headers *41*
12.2.3 Display Titles *41*
12.2.4 Column Headers *41*
12.2.5 Body of the Data Display *41*
12.2.6 Footnotes *43*

13 QUALITY CONTROL **45**

13.1 SPECIFICATIONS 45
13.2 OUTPUTS 45

14 APPENDICES **46**

14.1 CHANGES TO THE PROTOCOL SPECIFIED ANALYSES 46
14.2 SAMPLE SIZE RE-ESTIMATION PLAN 47

LIST OF ABBREVIATIONS

Abbreviation	Description
ACM	Ambulatory Cough Monitor
AE	Adverse Event
AICC	Corrected Akaike's Information Criterion
ANCOVA	Analysis of Covariance
AR(1)	1st Order Autocorrelation
BP	Blood Pressure
eCRF	Electronic Case Report Form
BDRM	Blinded Data Review Meeting
BLLQ	Below Lower Limit of Quantification
CIs	Confidence Intervals
CRO	Contract Research Organization
CS	Compound Symmetry
DLCO	Diffusion Capacity for Carbon Monoxide
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
ERS	Exposure-Response Set
FAS	Full Analysis Set
FEV_1	Forced Expiratory Volume in 1 Second
FVC	Forced Vital Capacity
GCP	Good Clinical Practice
GRC	Global Rating of Change
GRS	Global Rating of Status
H0	Null Hypothesis
H1	Alternative Hypothesis
HADS	Hospital Anxiety and Depression Scale
IB	Investigators Brochure
ICH	International Council for Harmonization
IEC	Independent Ethics Committee
IMP	Investigational Medicinal Product

Abbreviation	Description
IRB	Institutional Review Board
IPF	Idiopathic pulmonary fibrosis
IWRS	Interactive Web Response Services
K-BILD	Kings Brief Interstitial Lung Disease (questionnaire)
LCQ	Leicester Cough Questionnaire
LLOQ	Lower Limit of Quantification
MeDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
NK	Neurokinin
NRS	Numerical Rating Scale
PK	Pharmacokinetic
Pop PK	Population PK
PPS	Per Protocol Set
PRO	Patient Reported Outcome
PROMIS SD F 8b	Patient-Reported Outcomes Measurement Information System Sleep Disturbance Short Form 8b Questionnaire
RUCC	Refractory or Unexplained Chronic Cough
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SOC	System, Organ, Class
SOP	Standard Operating Procedure
SP	Substance P
SS	Safety Set
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TMF	Trial Master File
UN	Unstructured
VAS	Visual Analogue Scale

1 PURPOSE

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for NeRRe Therapeutics Study ORV-PF-01 (IPF COMFORT) and is based on protocol Versions 3.0/3.1, dates 05 April 2023[1].

The purpose of this SAP is to ensure that the data listings, summary tables and figures which will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

In the event of future amendments to the protocol, this SAP may need to be modified as necessary to account for changes relevant to the statistical analysis.

1.1 RESPONSIBLITIES

[redacted] will perform the statistical analyses and is responsible for the production and quality control of all derived datasets and tables, figures and listings.

2 INTRODUCTION

Orvepitant is a potent and selective inhibitor of the human neurokinin (NK) 1 receptor where it blocks the effects of Substance P (SP), the preferred endogenous ligand of the NK_1 receptor. The safety and efficacy of orvepitant has previously been evaluated in more than 900 patients and healthy volunteers, including two studies in patients with chronic cough. A full summary of the pharmacology, pre-clinical safety, pharmacokinetics and clinical safety and efficacy can be found in the current version of the Orvepitant Investigator's Brochure.

Idiopathic pulmonary fibrosis (IPF) is a rare and specific form of chronic, progressive, fibrosing interstitial pneumonia in which fibrosis is limited to the lungs. The prevalence is difficult to determine accurately because of the diagnostic criteria used and the populations examined, but the best estimates put the prevalence in the United States (US) at between approximately 14 per 100,000 using narrow criteria and 43 per 100,000 using broad criteria.

The incidence of IPF increases with older age, with presentation typically consisting of insidious onset of symptoms in the sixth and seventh decades of life. It is more common in men than women and most patients are current or past smokers. A history of gastro-oesophageal reflux is also common. IPF usually assumes a course of relentless physiologic deterioration although in some patients it remains stable for extended periods.

The symptoms of IPF have typically been present for months to years before diagnosis. Breathlessness with exertion and cough are the most common symptoms that lead to diagnosis. Fatigue and anxiety are also common symptoms. Bi-basilar fine inspiratory crackles (so-called Velcro crackles) are the most frequent physical examination finding and digital clubbing is seen in 25 to 50% of patients.

[1] Protocol version 3.0 (US) is active in the US and version 3.1 is active in the UK and Netherlands. In almost all respects 3.0 and 3.1 are identical save for minor country-specific changes in the UK and Netherlands versions. These differences do not affect the SAP.

The ORV-PF-01 study is being undertaken to evaluate the efficacy and safety of orvepitant as a treatment for the chronic cough associated with IPF.

2.1 STUDY OBJECTIVES

2.1.1 Primary Objective(s)

- To evaluate the effect of orvepitant once daily on cough severity, as perceived by patients, with IPF
- To evaluate the safety of orvepitant once daily in patients with IPF

2.1.2 Secondary Objective(s)

- To evaluate the effect of orvepitant once daily on other measures of cough burden and on health-related quality of life in patients with IPF
- To evaluate the effect of orvepitant on other comorbidities in patients with IPF

2.1.3 Exploratory Objective(s)

- To evaluate the effect of orvepitant once daily on markers of disease activity in patients with IPF
- To evaluate the relationship between plasma concentrations of orvepitant and efficacy in patients with IPF

2.2 STATISTICAL HYPOTHESIS

The statistical hypotheses for the primary efficacy endpoints are as follows:

Null hypothesis (H0): There is no difference in the mean change in mean cough severity NRS scores for the orvepitant treatment group compared to placebo.

H0: μ (active) = μ (placebo)

Alternative hypothesis (H1): There is a difference in the mean change in mean cough severity NRS scores for the orvepitant treatment group compared to placebo.

H1: μ (active) $\neq$ μ (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 (two in total: 10 mg versus placebo, 30 mg versus placebo) as each dose cohort is tested independently (versus the corresponding placebo) and each is adequately powered for 90% assuming a two-sided type I error of 0.05, not adjusted for multiple comparisons.

2.3 STUDY POPULATION

The study will include patients with IPF and troublesome chronic cough in whom the cough is attributed to the IPF.

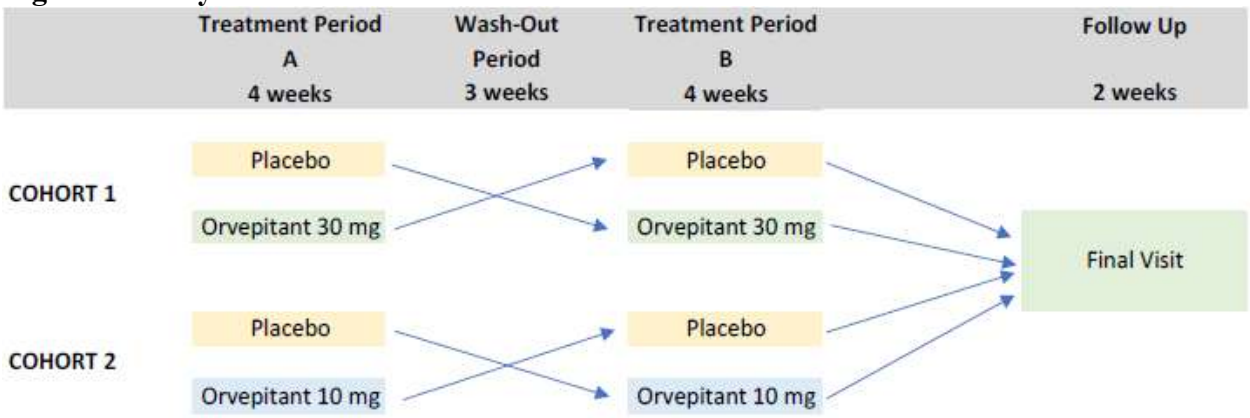
2.4 TRIAL DESCRIPTION

The study will be a multi-center, double-blind, randomised, placebo-controlled 2-period cross-over design. Two doses of orvepitant, 10 and 30 mg once daily, will be evaluated.

Subjects will participate in one of two cohorts (Cohort 1 [30 mg] and Cohort 2 [10 mg]). Within each cohort, subjects will be randomised to receive either orvepitant or placebo in the first treatment period (Treatment Period A) followed by the alternate treatment in Treatment Period B. There will be a wash-out period of 3 weeks between the two treatment periods. Subjects will be randomised 1:1 to each of the two treatment orders and 1:1 to each cohort.

Subjects will enter a screening period of up to 28 days to determine eligibility. Eligible subjects will be randomised at the Baseline visit and will participate in two identical 4-week treatment periods with the wash-out period between them. There will be a total of 8 visits including the Screening, Baseline and final Follow-up visits (Table 2 of the protocol).

Figure 1. Study Schematic



2.5 STUDY SAMPLE SIZE DETERMINATION

2.5.1 Estimated Sample Size

Assuming a common standard deviation of 2.7 for the difference between treatment periods, a sample size of 37 subjects per cohort has 90% power if the difference between treatments on the cough severity NRS is 1.5 points and assuming a two-sided type I error of 0.05, not adjusted for multiple comparisons. To allow for non-evaluable subjects/missing data the withdrawal rate and extent of missing data will be monitored as the trial progresses and the actual sample size will be increased accordingly such that data is available for 37 evaluable subjects.

As there have been no previous studies using NRS for cough severity the standard deviation of 2.7 was obtained based on historical data of the change in cough severity measured using VAS[2]. A 2-week observational study has since been completed in which the standard deviation of the change from baseline in NRS was 1.45 (95% CI: 1.14 - 2.01)[3]. Using the upper bound of the 95% CI and applying an inflation factor of 1.3 (derived from the observed increase in standard deviation from week 2 to week 12 in the historical VAS study data) resulted in an estimated standard deviation of 2.6. As this is very similar to, and lower than, the standard deviation originally used in the sample size estimation the estimated sample size required was not updated.

The proposed clinically relevant improvement of 1.5 points in the IPF coughing severity scale is an initial estimate based on the findings from the qualitative interviews with IPF patients conducted during the observational study. The qualitative interviews and psychometric analyses being undertaken as part of the current study will confirm (or else amend) this change for use in future studies.

2.5.2 Sample Size Re-Estimation

After 24 subjects have completed Treatment Period B (approximately 12 subjects per cohort), the variance of the cough severity NRS data will be reviewed on a fully blinded basis[4]. Up to a maximum of 20 subjects (approximately 10 per cohort) may be additionally recruited if the emerging cough severity NRS data is found to have greater variance than allowed for in the original sample size estimate. Any possible impact on the type I error due to this blinded sample size re-assessment procedure is believed to be negligible.

The Sample Size Re-Estimation Plan is provided in Section 14.2.

2.6 TREATMENT ASSIGNMENT AND BLINDING

2.6.1 Randomisation

Subjects will be randomised on Day 0 in a 1:1 ratio to either Cohort 1 or Cohort 2, and then within each cohort will be further randomised in a 1:1 ratio to receive either placebo or orvepitant in Treatment Period A. Randomisation will be performed centrally via an IWRS. The central randomisation list will be generated by [redacted] using SAS 9.4 (SAS Institute Inc., Cary, North Carolina, United States of America) or higher.

[2] VOLCANO-2 (A double blind, randomised, placebo controlled study of the efficacy and safety of three doses of orvepitant in subjects with chronic refractory cough), clinical study report, 09 Apr 2020.

[3] IPF-Obs-01 (An observational study to record cough symptom severity in patients with chronic cough due to idiopathic pulmonary fibrosis), clinical study report, 20 Sep 2021.

[4] The sample size re-estimate has been completed using data from 25 subjects and showed that the variance was not greater than that assumed variance in the original calculation. The required number of evaluable subjects per cohort therefore remains 37.

2.6.2 Blinding

The study is fully blinded and patients, site study staff or the study operations team will be unaware of both the cohort (and hence dose) to which a subject is randomised and the order of treatment. Blinding will be maintained until all subjects have completed the study and the database has been locked.

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 a subject's future medical care.

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

2.6.3 Planned and Actual Treatment

Planned treatment will be the one assigned per the randomisation list.

Actual (received) treatment might be different if the study medication kit(s) actually dispensed does not correspond to the kit(s) allocated by IWRS. Thus, actual treatment groups will be used for patients belonging to the Safety Analysis Set – according to the actual treatment dispensed to the patient. 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.

For the analysis of efficacy in the Full Analysis Set (FAS, study populations are defined in Section 4) the intent to treat principle will be followed and subject data will be assigned to the planned treatment group, regardless of the actual treatment provided.

Any occurrence of incorrectly administered treatment should be recorded as a protocol deviation.

2.6.4 Carry Over Effects

A potential limitation of any cross-over study is the occurrence of confounding carry over effects between treatment periods. The wash-out period for this study is 3-weeks which, given that the half-life of orvepitant is approximately 33 hours, is expected to be sufficient for the drug to be fully cleared and for cough symptoms to return to baseline levels.

Period-specific baseline values will be used in the analysis of primary and secondary efficacy endpoints, and safety endpoints, when available, to protect against possible carryover effect. Please refer to Section 5.2.2 for more details.

2.7 ADMINISTRATION OF TRIAL MEDICATION

IMP will be taken once daily during both parts of the study. IMP doses are fixed and will not be adjusted for individual subjects during the study.

At each visit (where appropriate, see Section 8.7 of the protocol) subjects will be dispensed IMP for self-administration for daily dosing until the subsequent visit, with overage included to allow for visit windows. An interim check of compliance will be made at all study visits (as appropriate).

Subjects will be required to take one tablet of IMP once-daily in the evening before retiring to bed. The study medication may be taken with or without food.

2.7.1 Missed dose

All doses are to be taken at approximately the same time each evening. If a subject forgets to take a dose in the evening it may be taken any time up until 9 am the following morning. After this time the dose should not be taken and it will be considered a missed dose.

2.7.2 Dose interruption

If 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 1 week is permitted. If on reintroduction of the IMP the AE recurs and remains intolerable the IMP will then be withdrawn altogether. A break in dosing will not result in extension of the overall dosing period.

2.8 SCHEDULE OF ASSESSMENTS AND PROCEDURES

Table 1: Schedule of Events

	Screening Visit	Baseline	Treatment Period A		End of Wash-Out / Start of Period B	Treatment Period B		Follow-up/ Early Termination[b]
			Week 2[f]	Week 4		Week 2[f]	Week 4	
Visit (V) Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Timing of visit	Within 14 to 28 days of V2	Day 0	14 days after V2	28 days after V2	21 days after V4	14 days after V5	28 days after V5	14 days after V7
Allowable window	N/A	0	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days
Informed Consent[a]	X							
Medical History & demographics	X							
Physical examination	X	X[c]						X[c]
Pulmonary function tests (spirometry)	X[j]	X		X	X		X	X
Inclusion/exclusion criteria	X	X[d]						
Concomitant medication review	X----------							----------X
Vital signs[e]	X	X		X	X		X	X
12-lead ECG	X	X		X	X		X	X
AE recording		X----------						----------X
Issue eDiary/Training/Compliance check	X	X	X	X	X	X		
Randomisation		X						
eDiary completion (once daily)[i]	X----------							----------X
Study drug dispensing/training		X			X			
Study drug collection/compliance check				X			X	
Ambulatory cough monitoring		X		X			X	
Daily dosing (evening before bed)			X[h] ----------	X		X[h] ----------	X	
Leicester Cough Questionnaire		X		X	X		X	
Hull Airway Reflex Questionnaire		X		X			X	
K-BILD Questionnaire		X		X	X		X	
PROMIS SF SD 8b		X		X	X		X	
Hospital Anxiety and Depression Scale		X		X	X		X	
Global ratings of status/change in cough		X[g]	X	X	X	X	X	X
Biomarker samples		X		X	X		X	
Clinical chemistry and haematology	X	X		X	X		X	X
Blood sample for orvepitant concentration				X	X		X	
Urinalysis	X	X		X	X		X	X

a	Informed consent will be obtained prior to any screening procedures being performed. Consent can be obtained prior to the Screening Visit or at the actual visit.
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 Follow-up visit and should be scheduled within 14 days of the early termination date or discontinuation of investigational product.
c	Symptom directed examination, if required.
d	Any changes that may impact eligibility are reviewed
e	Systolic and diastolic blood pressure, pulse rate, temperature, arterial oxygen saturation (measured using a pulse oximeter) and weight (visits 2 and 8 only for weight)
f	These visits will be conducted remotely/virtually (by phone or videoconference)
g	At baseline only the global rating of status is recorded
h	Starting the evening of the day following the Baseline and End of Wash-out Visits (to allow for completion of ambulatory cough monitoring at baseline). Finishing the evening of the Week 4 visit.
i	Used to record IPF Coughing Severity Scale, additional cough scales and other symptom scale scores. Last day of completion is the evening of Visit 7.
j	At screening, DLCO also measured if required

3 ENDPOINTS

3.1 PRIMARY EFFICACY ENDPOINT AND ESTIMAND

The primary endpoint is the mean change from baseline to Week 4 (the last 7 days of treatment) in weekly average of the daily IPF coughing severity scale score. The coughing severity scale is an 11-point numerical rating scale (NRS) that is completed daily by patients in an electronic diary.

3.1.1 ESTIMAND

3.1.1.1 Primary Estimand

The primary estimand includes the following attributes:

- Treatment of interest: Orvepitant (10 mg or 30 mg once daily) compared to placebo.
- Population of interest: Patients with IPF and troublesome chronic cough in whom the cough is attributed to the IPF, regardless of whether or not they receive prohibited medications during treatment or have other protocol violations.
- Variable / endpoint of interest: Change from baseline to Week 4 in weekly average (the average of the last 7 days before the relevant visit) of the daily IPF coughing severity scale score.
- Potential intercurrent events/Strategy to address:
    - Initiation of concomitant medications or withdrawal from study treatment: A treatment policy strategy will be used, essentially following the intention to treat principle. Any patients who discontinue treatment early will be followed up until the end of the study, with the aim of collecting all scheduled data.
    - Death during the study: A while on-treatment strategy will be used. All data for any subjects who died will be included up to the point of their death and no imputation will be made for any visits that were scheduled to have occurred after the date of death.
- Population-level summary: Difference in mean change from baseline to Week 4 in weekly average of the daily IPF coughing severity scale score between orvepitant (10 mg or 30 mg once daily) compared to placebo.

This estimand targets the effect of treatment on the variable measurement regardless of adherence to the treatment or the protocol. The primary analysis is to be done on the FAS and is expected to be sufficient for analysis of this estimand.

3.1.1.2 Supportive Estimand

The supportive estimand is cough severity in patients with IPF.

- Treatment of interest: Orvepitant (10 mg or 30 mg once daily) compared to placebo.

- Population of interest: Patients with IPF and troublesome chronic cough in whom the cough is attributed to the IPF, with no protocol violations which potentially confound the analysis of the primary efficacy endpoint.
- Variable / endpoint of interest: Change from baseline to Week 4 in weekly average (the average of last 7 days before the relevant visit) of the daily IPF coughing severity scale score.
- Potential intercurrent events: Subjects with intercurrent events that are considered protocol violations (e.g. prohibited concomitant medication) that may affect the primary efficacy analysis will have data excluded from the analysis. For other intercurrent events a similar strategy as for the primary estimand will be used.
- Population-level summary: Difference in mean change from baseline to Week 4 in weekly average of the daily IPF coughing severity scale score between orvepitant (10 mg or 30 mg once daily) compared to placebo.

This estimand targets the effect of treatment on the variable measurement in the target population in which intercurrent events which potentially confound the analysis of the primary efficacy endpoint did not occur. The primary analysis is to be repeated on the Per Protocol Set (PPS) and is expected to be sufficient for analysis of this estimand.

3.2 SECONDARY EFFICACY ENDPOINTS

The secondary efficacy endpoints comprise the following:

Assessment	Secondary Efficacy Endpoints
IPF coughing severity scale	• Mean change from baseline to Week 2 in weekly average of the daily IPF coughing severity scale
Early morning coughing scale	• Mean change from baseline to Weeks 2 and 4 in weekly average of the daily early morning coughing scale
Rest of the day coughing scale	• Mean change from baseline to Weeks 2 and 4 in weekly average of the daily rest of the day coughing scale
Urge to cough scale	• Mean change from baseline to Weeks 2 and 4 in weekly average of the daily urge to cough scale
Dyspnoea scale	• Mean change from baseline to Weeks 2 and 4 in weekly average of the daily dyspnoea scale
Cough frequency scale	• Mean change from baseline to Weeks 2 and 4 in weekly average of the daily cough frequency scale
Patient global ratings of status for all coughing, early morning	• Proportion of subjects in each category at Weeks 2 and 4

coughing and rest of the day coughing	
Patient global ratings of change for all coughing, early morning coughing and rest of the day coughing	• Proportion of subjects in each category at Weeks 2 and 4
Cough frequency measured using the Leicester Cough Monitor ambulatory cough monitor	• Mean change from baseline to Week 4 in 24-hour cough frequency • Mean change from baseline to Week 4 in awake cough frequency • Mean change from baseline to Week 4 in night-time cough frequency • Mean change from baseline to Week 4 in the number of coughing bouts
Leicester Cough Questionnaire (LCQ)	• Mean change from baseline to Week 4 in LCQ total and domain (Physical, Social, Psychological) scores
King's Brief Interstitial Lung Disease health status questionnaire (K-BILD)	• Mean change from baseline to Week 4 in K-BILD total and domain (Psychological, Breathlessness and Chest Symptoms) scores • Proportion of patients with a clinically relevant improvement in total K-BILD score
PROMIS SF SD 8b sleep assessment questionnaire	• Mean change from baseline to Week 4 in the PROMIS SF SD 8b score
Hospital Anxiety and Depression Scale (HADS) questionnaire	• Mean change from baseline to Week 4 in the HADS score
Hull Airway Reflux Questionnaire (HARQ)	• Mean change from baseline to Week 4 in HARQ score

3.3 EXPLORATORY ENDPOINTS

The following exploratory endpoints will be assessed:

Assessment	**Endpoint**
Plasma/serum biomarkers	• Mean change from baseline in the concentration of plasma/serum markers of inflammation and fibrosis

Plasma concentrations of orvepitant measured in spot samples	• PK exposure-response relationship for the orvepitant group will be carried out to examine the possible relationship between clinical efficacy and plasma levels of the drug

3.4 SAFETY ENDPOINTS

The safety endpoints comprise the following:

Assessments	**Safety endpoints**
Pulmonary function tests measured using a spirometer	• Change from baseline to each post-baseline assessment time-point in forced vital capacity (FVC), forced expired volume in one second (FEV1), peak expiratory flow rate and vital capacity (VC)
Adverse events recorded throughout the study	• 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	• Treatment emergent changes in physical examination will be recorded as adverse events
Vital Signs (pulse rate, systolic and diastolic blood pressure, arterial oxygen saturation, temperature, weight).	• Mean change from baseline to each post-baseline assessment time-point in each vital sign: - Systolic blood pressure - Diastolic blood pressure - Pulse rate - Temperature - Arterial oxygen saturation - Weight • Clinically significant changes in vital signs are recorded as adverse events

Haematology, clinical chemistry and urinalysis parameters	• Mean change from baseline to each post-baseline assessment time-point in each haematology parameter • Mean change from baseline to each post-baseline assessment time-point in each clinical chemistry parameter • Shift from baseline to each post-baseline assessment time-point in each urinalysis parameter • Clinically significant changes in laboratory parameters are recorded as adverse events
12-lead ECG parameters	• Proportion of subjects with clinically significant abnormal ECG findings at each post-baseline assessment time-point • Proportion of subjects with non-significant abnormal findings at each post-baseline assessment time-point • Change from baseline to each post-baseline assessment time-point in ECG intervals: - PR - QT, QTc and QTcF - RR • Proportion of subjects at each post-baseline assessment time-point with maximum absolute QTcF values by category - ≤450, >450 to ≤480, >480 to ≤500, >500 msec • Proportion of subjects at each post-baseline assessment time-point with maximum change from baseline in QTcF values by category - ≤0, >0 to ≤30, >30 to ≤60, >60 msec • Clinically significant changes in ECG findings are recorded as adverse events

4 ANALYSIS POPULATIONS

4.1 SAFETY SET (SS)

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.

4.2 FULL ANALYSIS SET (FAS)

All randomised subjects who received at least one dose of double-blind study drug, irrespective of treatment received, and have sufficient IPF coughing severity scale data for at least one post-treatment assessment. Subjects will be analysed according to randomised treatment. This set will be used for the primary efficacy analysis.

4.3 PER PROTOCOL SET (PPS)

All subjects in the FAS excluding those identified as having a relevant protocol violation, where a relevant violation is one that potentially confounds the analysis of the primary efficacy endpoint.

In addition, subjects who experience a respiratory adverse event that may result in a change in coughing status and that coincides with one or other of the evaluation periods will be excluded from the PPS.

Analysis Sets will be agreed at a blinded data review meeting (BDRM) prior to the unblinding of the study.

4.4 PROTOCOL DEVIATIONS

Protocol deviations must be documented in the subject's medical notes, in the electronic case report form, and/or [redacted] clinical trial management system.

Prior to database lock, NeRRe will be responsible for producing the final protocol deviation listing (formatted as a Microsoft Excel file), in collaboration with [redacted] The listing will be by subject and will include a description of the protocol deviation/violation, the occurrence date, the categorisation as major / minor and the categorisation as relevant / not relevant. Deviations classified as relevant will result in exclusion of the subject from the PPS. This file will be finalised and signed prior to database lock, at a BDRM.

A listing of all protocol deviations by subject and deviation category will be provided, indicating which are relevant deviations as determined before unblinding.

5 GENERAL ASPECTS FOR STATISTICAL ANALYSES

5.1 GENERAL METHODS

- All analyses and summaries will be produced using SAS® version 9.4 (or higher).
- All analysis will be analysed by cohort such that there will be two placebo groups presented unless specified otherwise.
- Categorical variables will be summarised using the number of observations (n), frequency and percentage of subjects. All percentages will be presented as one-decimal point, unless otherwise specified. Percentages equal to 100 will be presented as 100% and percentages will not be presented for zero frequencies.
- If log transformation is to be applied prior to analysis, a positive constant (1) will be added in cases of zero value prior to transformation. Analysis results will be back transformed to the original scale for reporting.
- Continuous variables and ordered categorical data will be summarised using the number of subjects with evaluable data, mean, standard deviation (SD), geometric mean (where

appropriate), median, first and third quartiles, minimum and maximum. In case of 0 values where geometric mean is calculated, 1 is added to every number in the data set, and then subtracted from the resulting geometric mean. The same number of decimal places as in the raw data will be presented when reporting minimum and maximum, 1 more decimal place than in the raw data will be presented when reporting the mean and median, and 2 more decimal places than in the raw data will be presented when reporting the SD. For ordered categorical data and nominal data, absolute counts and relative frequencies (in %) will be calculated.

- Lower and upper bound values for the confidence intervals (CIs) will be reported to 2 more decimal places than the raw data.
- All statistical hypothesis tests and CIs will be two sided, using a type I error rate of 0.05.
- Unless stated otherwise, the percentages will be based on the number of non-missing observations. The column header will still contain the number of subjects in the treatment group. There will be a row for the number of non-missing observations in the table (at each time point, if required) for each variable being summarised.
- Any calculated p-values will be presented to 3 decimal places; p-values less than 0.001 will be presented as ‘p<0.001’ and p-values greater than 0.999 will be presented as ‘p>0.999’.
- All relevant subject data will be included in listings and sorted by treatment group, Subject ID, and visit, as applicable, for all randomised subjects.
- Unscheduled or repeat assessments will not be included in summary tables unless specified. All assessments will be included in the subject listings.
- All tables, listings and figures will include footers that identify the name of the program that created the output, together with the date and time on which it was created. Headers will include the total number of pages that the presentation contains and, for each page, the number of the page within the presentation.

5.2 KEY DEFINITIONS

5.2.1 Study Day

Study Day 0 is the randomisation day. Subsequent days are numbered consecutively (Day 1, Day 2, etc.). Before Day 0, study days are numbered sequentially with negative values (i.e., Day -1, Day -2, etc.). This notation is used because the first dose of study medication is taken the day after the randomisation visit in Treatment Period A (and the day after Visit 5 in period B).

5.2.2 Baseline Values

Baseline values will be defined separately for each treatment period, unless stated otherwise.

For the primary efficacy endpoint (mean change from baseline to Week 4 in weekly average of the daily IPF Coughing Severity Scale score), period-specific baseline will be used in the analysis, which is the average of the 7 calendar days up to and including the date of Visit 2 for Treatment Period A and the average of 7 calendar days up to and including the date Visit 5 for Treatment Period B.

For secondary efficacy endpoints related to change from baseline, the period-specific baseline will be used when available; for those endpoints for which a period-specific baseline is not available

(cough frequency and Hull Airway Reflux Questionnaire), the baseline in Treatment Period A will be used.

For safety assessments, baseline will be Visit 2 for Treatment Period A and for Treatment Period B it will be Visit 5.

For both safety and efficacy assessments, if the period-specific baseline is missing for Treatment Period A the value from Visit 1 will be used. If this is missing change from baseline will not be calculated. If the period-specific baseline is missing for Treatment Period B the value from Visit 2 (first preference) or visit 1 (second preference) will be used. If both are missing change from baseline will not be calculated.

5.3 DATA HANDLING CONVENTIONS

5.3.1 Missing Data

The primary and several secondary efficacy endpoints are derived from data recorded by the study participants in the eDiary each day. To minimise missed completion, and so missing data, the compliance with eDiary completion is monitored by the sponsor daily. If a participant has missed entries, and in particular during the critical periods before study visits 2, 4, 5 and 7, the site staff are contacted and they in turn contact the patient to enquire about potential problems and remind them to complete the diary each day.

The secondary efficacy endpoint questionnaires that are completed periodically at the clinic visits are done under the direct supervision of the site staff. Each questionnaire has to be answered in its entirety before the next one becomes available on the iPad on which they are completed. The sites also have spare iPads in case of technical problems.

Collectively, these processes result in a high degree of compliance and hence minimal missing data for each of the PRO endpoints.

For the participants who are discontinued or withdraw from the trial, all data until the point of discontinuation and/or withdrawal from the study will be used in the summary and analyses. In general, missing values will not be imputed for the analysis, unless otherwise specified.

See Section 6.4 for imputation of partial dates for prior and concomitant medications.

See Section 8.1 for imputation of partial dates for adverse events.

See Section 7.1.1 for handling rules of missing data when deriving weekly average of NRS score.

5.3.2 Visit Windowing

For summaries of by visit safety data, the CRF visit will be used. Unscheduled visits will be listed but not mapped to a specific analysis visit.

For summaries and analyses of efficacy data, the following table outlines the analysis visits that efficacy data will be mapped to. Efficacy assessments will be mapped to a study analysis visit as long as the study day(s) falls within the windows specified in Column A (Protocol allowed window for the visit) or Column B (Reasonable limits for out of window days for actual visit to be associated with nominal Visit). Efficacy data collected on days falling out of analysis visit windows (Column C) will be reviewed at the BDRM and may be assigned to an analysis visit window if appropriate.

Study Analysis Visit	**Target Day for CRF Visit**	**A. Protocol allowed window for the visit (inclusive)**	**B. Reasonable limits for out of window days for actual Visit to be associated with nominal Visit**	**C. Unaccounted for days**
Baseline (Visit 2)	0	N/A	N/A	N/A
Treatment Period A - Week 2 (Visit 3)	14	11 to 17	7[a] to 10 18 - 20	1 to 6
Treatment Period A - Week 4 (Visit 4)	28	25 to 31	21 to 24 32 to 34[b]	35 to 39
[Wash-out start]	29	N/A	N/A	N/A
End of Wash-Out / Start of Period B (Visit 5)	B0[c] [49]	N/A [46 to 52]	N/A [42 to 55]	N/A
Treatment Period B - Week 2 (Visit 6)	B14 [63]	B11 to B17 [60 to 66]	B7[a] to B10 [56 to 59] B18 to B20 [67 to 69]	B1 to B6
Treatment Period B - Week 4 (Visit 7)	B28 [77]	B25 to B31 [74 to 80]	B21 to B24 [70 to 73] B32 to B34[b] [81 to 83]	B35 to B39

a. Earliest day on which a 7 day data collection period is possible
b. Last day a fully compliant subject will be able to dose (subjects are provided with 34 days of study medication)
c. Please note that for the purposes of visit window assignment only the clock is re-set for treatment period B
[ ] = nominal study day, where Baseline Visit 2 = Day 0; BDRM = blinded data review meeting

The weekly average score is derived using scores recorded in the eDiary from the last 7 calendar days up to and including the date of the relevant visit (as defined in the visit window above).

At least 4 scores are required to derive a weekly average. These four scores do not have to be on consecutive days. If 4 scores are not recorded in the eDiary the weekly average will be considered missing for that visit.

5.3.3 Data Dates

For diary data in particular, it is possible that the end date of the task can fall onto the day after the start date of the task (for example, if the diary completion is started just before midnight). In such instances, the start date will be used as the assessment date.

If the patient-reported date of first dose in each treatment period is missing, it will be assumed that the first dose is taken the day after the date of dispensing the drug (i.e., the day after the date of Visit 2 for Treatment Period A and the day after the date of Visit 5 for Treatment Period B).

5.3.4 Pooling of Sites

All sites will be pooled together for analysis purposes.

6 STUDY SAMPLE

Summary tables for disposition, demographics, baseline characteristics, medical history, prior and concomitant medications will be summarised by dose group.

6.1 DISPOSITION OF SUBJECTS

Summary statistics will tabulate the number and percentage of subjects who are screened, screen failures, randomised, received at least one dose of study treatment, who completed the study and who terminated the study early along with associated reasons. This table will also show the number and percentage of subjects who completed the first treatment period, who completed second treatment period, and who withdraw from treatment from each period along with associated reasons. The number and percentage of subjects included in each of the analysis populations will be presented. No statistical testing will be performed on these data. The number of randomised subjects in each dose group will be used as the denominator to calculate percentages. The disposition table will be based on all screened subjects.

Subject disposition data listings will be presented.

6.2 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographics (age, race, ethnicity, sex, weight (kg)) and baseline characteristics will be summarised by each dose group and overall.

The demographic and baseline characteristics table will be based on the SAS. Subject demographic data listings will be presented.

6.3 MEDICAL HISTORY

Medical history will be coded using the latest available version of the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary (version 26.0 [March 2023] or higher). Summary tables of the number and percentage of subjects with medical history by system organ

class (SOC) and preferred term (PT) will be produced for subjects in the SAS. Medical history will be sorted by SOC and PT in descending order of frequency in the overall column (in case of a tie, alphabetical order will be followed). For the summary tables, a subject may appear more than once if they have more than one medical history finding coded under different SOC terms, or more than one medical history finding with a different PT under the same SOC term. However, the subject will be counted only once in the overall category (SOC).

Subject medical history listings with coded SOC and PT along with verbatim eCRF term will also be provided.

6.4 PRIOR AND CONCOMITANT MEDICATIONS

Prior medications are defined as medications that started and stopped before the date of first dosing with study medication. Any medication that started on the date of dosing will not be considered prior. Any medication that stopped on the date of dosing will be considered prior.

Concomitant medications are defined as all medications (excluding study treatment) started or ongoing on/after the date of first dosing with study drug. Medications started or ongoing on/after the date of first dosing in Period A or during the wash-out period will be considered as concomitant medications for Period A; Medications started or ongoing on/after the date of first dosing in Period B up to end of follow-up will be considered as concomitant medications for Period B. Partial start dates in prior and concomitant medications will be imputed with a conservative algorithm, to the first day of the month (if missing day) or the first month of the year (if missing month). Partial end dates in prior and concomitant medications will be imputed to the last day of the month (if missing day) or the last month of the year (if missing month). If the year is missing from the start date the medication will be considered to have been started before the date of Visit 1 for that subject.

Prior and concomitant medications will be coded by using the latest available Anatomical Therapeutic Chemical (ATC) Level 2 classification code and preferred drug name according to the World Health Organization Drug Dictionary (WHODrug) Global (March 2023 or higher). Prior and concomitant medications will be sorted by subject ID and then by start date.

Separate summary tables will be provided for prior and concomitant medications, presenting the number and percentage of subjects by each dose group; the summaries will be sorted by ATC Level 2 and PT in descending order of frequency in the overall column (in case of a tie, alphabetical order will be followed). For each subject, the medication will be counted only once within a given ATC level 2 and only once within a given preferred drug name level. A subject may appear more than once if he/she has more than 1 concomitant medication coded under different ATC categories; however, the subject will be counted only once in the overall category.

All medications will be listed.

7 EFFICACY ANALYSES

All efficacy endpoints, including responder rates, will be summarised separately for each cohort with descriptive statistics by cohort and visit as appropriate.

7.1 PRIMARY AND SECONDARY EFFICACY ANALYSIS

Efficacy endpoints will be summarised separately for each cohort by descriptive statistics or counts and percentages for each treatment group at each time point, as appropriate. All statistical hypothesis tests and confidence intervals will be two sided, using a type I error rate of 0.05. No adjustments for multiple comparisons will be used.

The change from baseline of parameters in Week 4 (and Week 2 where also relevant) will be summarised for each treatment group in the FAS population.

Statistical comparisons for the primary and secondary endpoints will be done separately for each cohort as described below in each subsection. For the primary endpoint and secondary endpoints that are a daily numerical or verbal rating scale, an analysis of covariance (ANCOVA) model will be used with the response variable as the change from baseline in the mean score at the timepoint of interest, and predictor variables (fixed effects) as treatment, baseline mean score for the endpoint, period and interaction effect between treatment and period. Subject will be included as a random effect. Investigation of other covariates such as age and sex and their impact on the results may be done post-hoc. Sample SAS code for the ANCOVA is provided below.

```
proc mixed data= Test;
[redacted]
;
run;
```

Normality of error terms may be checked with QQ plots or other methods. If normality assumptions are considered severely violated, log transformation will be attempted first to see if that fixes the issue. If the normality assumption is still violated after applying transformation, a non-parametric method (e.g., Wilcoxon signed-rank test) will be used for conducting treatment comparison between orvepitant (10 mg or 30 mg once daily) and placebo.

All efficacy data, both patient-reported scores/assessments and derived parameters, will be listed.

7.1.1 Primary Efficacy Endpoint

7.1.1.1 Main Analysis of Primary Efficacy Endpoint

The primary endpoint is the mean change from baseline to Week 4 (the average of the last 7 days before the relevant visit) in weekly average of the daily IPF Coughing Severity Scale score.

- The weekly average score is derived using NRS scores recorded in the eDiary from the last 7 calendar days up to and including the date of the relevant visit (please also refer to Section 5.3.2 for details).
- At least 4 scores are required to derive a weekly average. These four scores do not have to be on consecutive days.
- If 4 scores are not recorded in the eDiary the weekly average will be considered missing.

Change from baseline to Week 4 in weekly average of the daily IPF Coughing Severity Scale score will be analyzed using the ANCOVA model as described in Section 7.1. Least square means and corresponding 95% confidence intervals (CIs) will be presented by treatment group/cohort. Mean difference, corresponding CIs and p-value for treatment comparison between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. The mean change from baseline and the corresponding 95% confidence intervals (CIs) of the primary endpoint from the ANCOVA analysis will be displayed graphically at Week 4 by treatment group/cohort.

7.1.1.2 Supportive/Sensitivity Analysis of Primary Efficacy Endpoint

Mixed Model for Repeated Measures (MMRM)

Change from baseline in weekly average of the daily IPF Coughing Severity Scale score at Week 2 and Week 4 will also be analysed using mixed model for repeated measures (MMRM). The model will include treatment, time point (Week 2 or Week 4), treatment by time point, period and treatment by period as fixed effects and subject as a random effect. Baseline will be included as a covariate. Sample SAS code for MMRM is provided below.

```
Proc mixed data= Test;
[redacted]
run;
```

As measurements over time within each period are expected to be correlated, a covariance structure will be used to model the correlation. Several covariance structures will be tested, including unstructured (UN), compound symmetry (CS), 1st order autocorrelation [AR(1)]. The covariance structures with the smallest Akaike's information criterion (AICC) will be chosen and used in the model.

Least square means and corresponding 95% CIs at Week 2 and Week 4 will be presented by treatment group/cohort. Mean difference, corresponding CIs and p-value for treatment comparison between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented at Week 2 and Week 4. The mean change from baseline and the corresponding 95% confidence interval (CI) will be displayed graphically by visit (Week 2 and Week 4) for the observed values and by treatment group/cohort.

Sensitivity Analyses

A sensitivity analysis will be done on the primary endpoint, repeating the primary analysis using the PPS which excludes those with protocol deviations that are judged to have a potential impact

on the analysis of the primary endpoint. This sensitivity analysis is expected to be sufficient to evaluate the supportive estimand discussed in Section 3.1.1.2.

As outlined in Section 7.1.1.1 the primary endpoint will be derived conditional on there being scores available on at least 4 of the 7 days over which the average will be calculated. To evaluate the potential impact of this rule for dealing with missing data in deriving the average score an additional analysis will be run which includes the weekly average, where the weekly average is derived regardless of the number of days data available within that week. That is, it will include a weekly average calculated for weeks in which there is at least one NRS score recorded. Depending on the level of missing data further derivations of the weekly average may be examined post-hoc.

7.1.2 Secondary Efficacy Endpoints

7.1.2.1 IPF Coughing Severity Scale

The mean change from baseline to Week 2 in weekly average of the daily IPF Coughing Severity Scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint at Week 2. Least square means and corresponding 95% CIs will be presented by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparisons between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. Derivation of weekly averages will use the same rules as for the primary endpoint.

7.1.2.2 Early Morning Coughing Scale

The mean change from baseline to Weeks 2 and 4 in weekly average of the daily early morning coughing scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint separately by time point (Week 2 or Week 4). Least square means and corresponding 95% CIs will be presented at Week 2 and Week 4 by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparisons between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. Derivation of weekly averages will use the same rules as for the primary endpoint.

7.1.2.3 Rest of The Day Coughing Scale

The mean change from baseline to Weeks 2 and 4 in weekly average of the daily rest of the day coughing scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint separately by time point (Week 2 or Week 4). Least square means and corresponding 95% CIs will be presented at Week 2 and Week 4 by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparisons between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. Derivation of weekly averages will use the same rules as for the primary endpoint.

7.1.2.4 Urge To Cough Scale

The mean change from baseline to Weeks 2 and 4 in weekly average of the daily urge to cough scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint separately by time point (Week 2 or Week 4). Least square means and corresponding 95% CIs will be presented at Week 2 and Week 4 by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparisons

between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. Derivation of weekly averages will use the same rules as for the primary endpoint.

7.1.2.5 Dyspnoea Scale

The mean change from baseline to Weeks 2 and 4 in weekly average of the daily dyspnoea scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint separately by time point (Week 2 or Week 4). Least square means and corresponding 95% CIs will be presented at Week 2 and Week 4 by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparisons between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. Derivation of weekly averages will use the same rules as for the primary endpoint.

7.1.2.6 Cough Frequency Scale

The mean change from baseline to Weeks 2 and 4 in weekly average of the daily cough frequency scale will be assessed. An ANCOVA model, similar to that suggested for the primary endpoint, will be used to assess this endpoint separately by time point (Week 2 or Week 4) and cohort. Least square means and corresponding 95% CIs will be presented at Week 2 and Week 4 by treatment group/cohort. Mean differences, corresponding CIs and p-values for treatment comparison between orvepitant (10 mg or 30 mg once daily) and placebo will also be presented. For these analyses, the daily verbal rating scale scores will be converted to a numerical values (where Never = 0 and All the time = 4) and derivation of weekly averages will use the same rules as for the primary endpoint.

In addition, a summary for the cough frequency scale will be presented of the modal response during the 7 days prior and up to each assessment time point. Frequency tables will display the frequency and percentage of subjects in each category at Weeks 2 and 4. Shift from baseline tables will also be provided at Weeks 2 and 4 with baseline category as columns and Week 2 (or Week 4) category as rows.

7.1.2.7 Patient Global Rating of Status

Patient global ratings of status for all coughing, early morning coughing and rest of the day coughing will be summarised with frequency tables by displaying the frequency and percentage of subjects in each category at Weeks 2 and 4. Shift from baseline tables will also be provided at Weeks 2 and 4 with baseline category as columns and Week 2 (or Week 4) category as rows.

7.1.2.8 Patient Global Rating of Change

Patient global ratings of change for all coughing, early morning coughing and rest of the day coughing will be summarised in the same way as the patient global rating of status with frequency tables for each category at weeks 2 and 4. Shift from baseline summaries will be not applicable as there is no baseline recorded for this endpoint.

Additionally, the patient global rating of change analysis will combine the numerical levels into the following groups for each question in the patient global rating analysis:

- Improved Global Rating of Change: Scores 1, 2, 3
- No Change: Score 0

- Worsened: Scores -3, -2, -1

For each question, frequency tables will be provided for each of the combined categories above at weeks 2 and 4.

7.1.2.9 Leicester Cough Questionnaire

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.

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 primary endpoint. That is, an ANCOVA model will be used to assess change from baseline in LCQ total score and separately for each domain from baseline to Week 4.

The Domains are comprised of the following 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

The individual domain scores are calculated by dividing the total score from items in the domain by the number of items in the domain.

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

The 10% rule, as recommended by the originator of the assessment, will be used for missing data as follows:

- 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.

7.1.2.10 Cough Frequency

Cough frequency is measured using the Leicester Cough Monitor ambulatory cough monitor applied for 24-hours at the Baseline visit (Period A) and in Week 4 of each treatment period.

Cough frequency data will be log transformed before modelling (descriptive summaries will still be performed using the original values). If there are subjects with cough frequencies of 0, 1 will be added prior to log transformation and will then be subtracted from the back-transformed value. Data will be tested for normality prior to transformation. An ANCOVA model similar to the primary endpoint will be used except only baseline for Period A will be used as there is no baseline available for Period B. The following will be assessed for Mean change from baseline to Week 4:

- Average number of coughs per hour in the 24 hour period
- Average number of coughs per hour in the awake period
- Average number of coughs per hour in the asleep period
- Average number of bouts per hour in the 24 hour period
- Average number of bouts per hour in the awake period
- Average number of bouts per hour in the asleep period
- Average number of coughs in a cough bout in the 24 hour period
- Average number of coughs in a cough bout in the awake period
- Average number of coughs in a cough bout in the asleep period.

The treatment effect 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 ratio of GM ratios is the geometric mean ratio (orvepitant [10 mg or 30 mg once daily]/ Placebo) of the Week 4/baseline result. The adjusted geometric mean (GM) ratio (Week 4/baseline) and corresponding 95% CIs for each treatment group, ratio of GM ratios (orvepitant [10 mg or 30 mg once daily]/Placebo), 95% CIs for the ratio of GM ratios and the p-value from the hypothesis test of no difference between the treatment groups at Week 4 will be presented by cohort.

7.1.2.11 King's Brief Interstitial Lung Disease Questionnaire

The K-BILD is a brief, validated, self-completed health status measure for interstitial lung diseases. It comprises 15 questions in three domains. Domain scores are calculated as logit-transformed weighted sums of individual items within each domain. Domains are Psychological (items 3, 5, 6, 8, 10, 12, 14), Breathlessness (1, 4, 11, 13) and Chest Symptoms (2, 7, 9). (Note that item 15 does not belong to any of the domains and only contributes to the total score.) Domain scores are not calculated if 50% or more of the individual items for that domain are missing. Total score is calculated as the logit-transformed weighted sum of all items. Total score is not calculated if any of the three domain scores is missing. K-BILD will be completed by subjects at Visits 2, 4, 5 and 7.

If an item is missing, the patient's average score for items available for a specific domain is used as the value for the missing item (but only if most items in domain are available).

The average score for items available for a specific domain is used as the value for the missing item (but only if >50% items in domain are available).

Mean change from baseline to week 4 in total and domain scores will be analysed with an ANCOVA model similar to that used for the primary endpoint.

Additionally, the proportion of patients with a clinically relevant improvement in K-BILD total score[5] and 95% Clopper–Pearson CIs will be presented at week 4. Fisher's Exact Test will be used for computing p-value for treatment comparison between orvepitant (10 mg or 30 mg once daily) and placebo if computationally feasible (otherwise, the Chi-Square Test may be used if the expected count in any cell is greater than 5).

7.1.2.12 PROMIS SF SD 8b Sleep Questionnaire

The PROMIS SD SF 8b is a short form questionnaire to assess sleep disturbance. It comprises eight questions and will be completed by subjects at Visits 2, 4, 5 and 7. Total score is calculated as sum of individual items. Total score is not calculated if any individual item is missing.

Mean change from baseline to week 4 for each question and the total score will be analysed with an ANCOVA model similar to that used for the primary endpoint.

7.1.2.13 Hospital Anxiety and Depression Scale Questionnaire

The HADS tool is used for assessing anxiety and depression which comprises of 14 questions, seven each for anxiety and depression. Total anxiety score is calculated as sum of the seven odd-numbered questions, total depression score as sum of even-numbered questions. Total scores are not calculated if more than one item from that domain is missing. If a single item is missing, then the patient's average score for items available for a specific score is used as the value for the missing item. HADS will be completed by subjects at Visits 2, 4, 5 and 7.

Mean change from baseline to week 4 for the sum of the anxiety and depression questions will, separately, be analysed with an ANCOVA model similar to that used for the primary endpoint separately by cohort.

7.1.2.14 Hull Airway Reflux Questionnaire

The Hull Airway Reflux Questionnaire is a 14-item questionnaire that assesses airway hypersensitivity, particularly that associated with reflux of gastric contents. Total score is calculated as sum of individual items. The questionnaire will be completed by subjects at Visits 2, 4 and 7.

Mean change from baseline to week 4 for the total score will be analysed with an ANCOVA model similar to that used for the primary endpoint separately by cohort except only baseline for Period A will be used as there is no baseline available for Period B.

7.1.3 Sub-Groups

The primary efficacy endpoint will also be summarised descriptively by the following subgroups, provided there are a minimum of 10 evaluable subjects in each sub-group within the category and cohort:

[5] Total score increase of ≥5.0 (et al. Sinha A, Patel AS, Siegert RJ, et alThe King's Brief Interstitial Lung Disease (KBILD) questionnaire: an updated minimal clinically important difference BMJ Open Respiratory Research 2019;6:e000363. doi: 10.1136/bmjresp-2018-000363)

Category	Sub-group
Use of concomitant disease modifying therapies (nintedanib or pirfenidone) at baseline	Yes (using) No (not using)
IPF Cough Severity Scale score at baseline	≤ Median > Median
Sex	Male Female
Race	White Non-white
Ethnicity	Hispanic Non-hispanic

The awake cough frequency endpoint will also be summarised descriptively in the following sub-group:

- By awake cough frequency at baseline (<=Median, >Median)

8 SAFETY ANALYSES

Safety data will be summarised descriptively for each treatment. That is, safety data will be summarised for two doses of orvepitant (10 mg and 30 mg separately), and two placebo groups separately by cohort.

No formal inferential tests will be performed on safety data.

Safety analysis will be carried out using the SS population.

8.1 ADVERSE EVENTS

Adverse events will be coded using the latest available version of MedDRA (v26.0 or higher). All treatment-emergent AEs (TEAEs) from the start of study drug dosing will be collected. A TEAE is any AE that first occurred or worsened in severity following first dose. AEs that start during the wash-out period will be assigned to the treatment received during Treatment Period A for TEAE reporting; AEs that start following first dose in Treatment Period B up to end of follow-up will be assigned to the treatment received during Treatment Period B for TEAE reporting.

The following AE summaries will be presented:

- An overall summary table, including the number of events and number and percentage of subjects with:
  - TEAEs
  - Subjects with at least one TEAE
  - Study drug-related TEAEs

  - Serious TEAEs
  - TEAEs with an outcome of death
  - Treatment-emergent AESIs
  - Severe TEAEs
  - TEAEs leading to study drug discontinuation

- TEAEs by SOC and PT. This table will have a section for overall, and two subgroup sections for concomitant use of disease modifying therapies (yes, no [nintedanib or pirfenidone])
- Serious TEAEs by SOC and PT
- Treatment-emergent AESIs by SOC and PT
- TEAEs leading to study drug discontinuation by SOC and PT
- Related TEAEs by SOC and PT
- Maximum severity of TEAEs by SOC and PT

Non-treatment emergent AEs and TEAEs will be listed.

In addition, the following subject listings will be provided as tables:

- Deaths
- Serious TEAEs
- AESIs
- TEAEs leading to study drug discontinuation

For all AE tables summarised by SOC and PT, a subject contributes only once to the count for a given AE on the SOC level and on the PT level within SOC. SOC and PT levels will be sorted in descending order of frequency (subject count), followed by descending order of frequency (event count), followed by alphabetical order.

AEs will be classified as either related or unrelated to study medication. AEs with missing relationship to study drug are counted as related in the summary tables.

In the summary by severity, subjects reporting AEs at different severities will be counted only once at the highest severity reported within an AE level (SOC or PT). Severity categories will comprise of mild, moderate and severe.

Partial start dates in AEs will be imputed with a conservative algorithm, to the first day of the month (if missing day) or the first month of the year (if missing month). Partial end dates in adverse events will be imputed to the last day of the month (if missing day) or the last month of the year (if missing month). All AEs will be assigned as a TEAE or not based on the imputed dates as appropriate. If the onset date is missing a conservative approach will be applied such that the AE is considered treatment emergent.

8.2 CLINICAL LABORATORY TESTS

Blood and urine samples will be obtained for assessment of clinical laboratory parameters. Clinical laboratory assessments (haematology, chemistry, urinalysis) will be performed on the visits as described in the schedule of study assessments and procedures in the protocol.

Clinical laboratory parameters at all scheduled visits will be summarised using appropriate descriptive summary statistics.

Period-specific baseline values will be used (see Section 5.2.2). The change from baseline in each haematology and chemistry to each post-baseline visit will also be summarised by treatment group.

Shifts from baseline for laboratory results categories (Low, within normal range, high) to each study visit will be presented by treatment group for each laboratory parameter.

All clinical laboratory results will also be listed. Laboratory values that are outside the normal range will be flagged in the participant data listing. Listings of abnormal laboratory parameters will also be provided.

8.3 VITAL SIGNS

Vital signs data recorded in the study include systolic and diastolic blood pressure, pulse rate, temperature and arterial oxygen saturation (SpO2). Descriptive statistics of vital signs will be presented by treatment group for all scheduled visits using appropriate descriptive summary statistics.

Period-specific baseline values will be used (see Section 5.2.2). The change from baseline to each post-baseline visit will also be summarised by treatment group.

Clinically significant abnormalities will be summarised using count and percentages for vital signs overall interpretation at all scheduled visits by treatment group and cohorts.

All vital signs data will be included in subject data listings, and clinically significant abnormalities will be flagged.

8.4 ECG DATA

ECGs will be assessed by the investigator for overall abnormal findings and will be classified as Normal, Abnormal not clinically significant or Abnormal clinically significant. These will be summarised by the proportion of subjects within each category at each post-baseline assessment by treatment group. Shift tables will also be presented.

The actual value and change from period-specific baseline (see Section 5.2.2) will be summarised descriptively by treatment group and time point for each ECG interval parameters by treatment group.

- ECG intervals:
    - QRS duration

- PR
- QT, QTc and QTcF
- RR

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

In addition, the proportion of subjects at each post baseline assessment with maximum absolute QTcF values by the following categories will be presented:

- ≤450 msec
- >450 to ≤480 msec
- >480 to ≤500 msec
- >500 msec

Similarly, the proportion of subjects at each post baseline assessment with increase from baseline in QTcF values by the following categories will be presented:

- ≤0 msec
- >0 to ≤30 msec
- >30 to ≤60 msec
- >60 msec

All ECG parameters will be listed.

8.5 PHYSICAL EXAMINATION

Physical examination data include measurements as collected in the CRF such as General appearance, Skin, Head, eyes, ears, nose and throat, Respiratory, Cardiovascular, Abdomen (including liver and kidneys), Musculoskeletal, and Neurological.

A by-subject listing with physical examination findings will be provided.

8.6 PULMONARY FUNCTION TESTS

Pulmonary function test (PFT) data recorded in the study include forced vital capacity, forced expired volume in 1 second and peak expiratory flow rate and vital capacity (VC). Descriptive statistics of the PFTs will be presented by treatment group for all scheduled visits and will be summarised using appropriate descriptive summary statistics by treatment group.

Period-specific baseline values will be used (see Section 5.2.2). The change from baseline to each post-baseline visit will also be summarised by treatment group.

Clinically significant abnormalities will be summarised using count and percentages for each parameter examined at all scheduled visits by treatment group.

All PFT data will be included in subject data listings.

8.7 TREATMENT COMPLIANCE AND EXPOSURE

Definitions of exposure variables are provided in Table 2.

Table 2: Exposure variables definitions

Variables	Definitions
Duration of dosing (days)	Last dose of treatment – first dose of randomised treatment +1 The first and last days of treatment in each treatment period is used to derive the duration of dosing for each treatment (placebo and orvepitant).
Study treatment compliance (%)	100 * (total number of doses taken) / (total number of doses planned to be taken) Total number of doses taken is calculated using the eCRF pages 'Drug Accountability" and "Drug Dispensing", based on the following: Each bottle contains 34 tablets which is sufficient for the full 28 day dosing period plus overage. Number of doses taken = number of tablets dispensed (at Visit 2 [Period A] or Visit 5 [Period B]) – number of tablets remaining in the bottle (at Visit 4 [Period A] or Visit 7 [Period B]). The total number of doses planned to be taken is the number of days between the first visit in the period (Visit 2 or 5) and the Week 4 visit (Visit 4 or 7) for each period. If a subject is early discontinued prior to Week 4 within the period, 28 will be used for the total number of doses planned within the period.
Cumulative orvepitant exposure (mg)	In placebo period, Cumulative dose (in mg) = 0 In orvepitant 10 mg period, Cumulative dose (in mg) = number of doses taken * 10 In orvepitant 30 mg periods, Cumulative dose (in mg) = number of doses taken * 30

Note, if a subject interrupts treatment (for example due to an AE) but subsequently restarts treatment and completes, in total, the planned 4-week treatment period, an adjustment to the actual date range will be applied to the compliance calculation to include only those days that the subject was expected to take treatment (i.e., the dates during which treatment was interrupted will be excluded from the calculation).

The following information will be tabulated by treatment:

- Summary of drug duration
- Summary of cumulative orvepitant exposure (in mg)
- Summary of randomised study treatment compliance: Compliance will be summarised quantitatively and qualitatively for the category (≤ 50%, >50 to ≤ 80%, >80 to <100%, =100%, >100 to ≤ 120%, >120%).

Listings of IMP dispense/administration and drug accountability data will be provided for the Safety Analysis Set.

9 BIOMARKERS

Plasma/serum concentrations of each biomarker (those specifically listed in section 10.3 of the protocol) at each study visit will be summarised descriptively by treatment group for each cohort. Mean (SD), minimum, maximum and median (quartile) for changes from baseline (period-specific; see Section 5.2.2) to each post-baseline visit will be summarised for each analyte. Values below the limit of quantification will be set to 0.

10 PHARMACOKINETIC ANALYSES

To explore the PK exposure-response relationship of two doses of orvepitant in the participants a Pop PK approach by nonlinear mixed effects modeling will be used in exposure-response set. Relationships will be explored graphically, using linear or smoothing functions to evaluate relationships. Further model based-analyses will be conducted, including non-linear mixed effects modelling (NONMEM®) of orvepitant to evaluate covariates and subsequent impact on biomarkers and clinical endpoints. The details of these analyses will be described in a separate Pop PK Data Analysis Plan and the data will be reported separately from the safety and efficacy findings.

Plasma concentrations of orvepitant will be summarised descriptively at the end of washout and week 4 of each treatment period. 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.

Separate summaries of orvepitant plasma concentrations will also be presented for the sub-groups taking pirfenidone and nintedanib concomitantly.

11 INTERIM ANALYSIS

There is no formal interim analysis planned at this time. The final database lock will take place when the last subject has completed the last visit for the study.

11.1 DATA SAFETY MONITORING BOARD (DSMB)

A DSMB will be convened with the primary remit of providing independent safety oversight of the study.

The composition of the DSMB, its terms of reference and its operating procedures are described in the DSMB Charter which will be finalised before the first subject is randomised into the study. The DSMB membership will include an unblinded statistician who will be responsible for providing the DSMB with the agreed information necessary for the DSMB to undertake its assessments of safety.

12 SOFTWARE AND PROGRAMMING SPECIFICATIONS

All datasets, TLFs, and statistical analyses will be generated using SAS, Release 9.4 or higher (SAS Institute Inc., Cary, NC, USA). Computer-generated datasets, table, listing and figure output will adhere to the following specifications:

12.1 GENERAL PROGRAMMING SPECIFICATIONS

- One SAS program can create several outputs, or a separate SAS program can be created for each output at the statistical programmer's discretion.
- Each output will be stored in a separate file.
- Output files will be delivered in Word format / rtf format.
- Numbering of TLFs will follow ICH E3 guidance

12.2 TABLE, LISTING, AND FIGURE FORMAT

12.2.1 General

- All TLFs will be produced in landscape format, unless otherwise specified.
- All TLFs will be produced using the Courier New font, size 8.
- The data displays for all TLFs will have a 1.5-inch binding margin on top of a landscape-orientated page and a minimum 1-inch margin on the other 3 sides.
- Headers and footers for figures will be in Courier New font, size 8.
- Legends will be used for all figures with more than 1 variable, group, or item displayed.
- TLFs will be in black and white (no color), unless otherwise specified.
- Specialized text styles, such as bolding, italics, borders, shading, and superscripted and subscripted text, will not be used in the TLFs, unless otherwise specified. On some occasions, superscripts 1, 2, or 3 may be used (e.g., cm^2, C_{max}).
- Only standard keyboard characters will be used in the TLFs. Special characters, such as non-printable control characters, printer-specific, or font-specific characters, will not be used. Hexadecimal-derived characters will be used, where possible, if they are appropriate to help display math symbols (e.g., μ). Certain subscripts and superscripts (e.g., cm^2, C_{max}) will be employed on a case-by-case basis.
- Mixed case will be used for all titles, footnotes, column headers, and programmer-supplied formats, as appropriate.

12.2.2 Headers

- All output should have the following header at the top left of each page:

  NeRRe Therapeutics Ltd
  Protocol: ORV-PF-01

- All outputs should have Page n of N at the top or bottom right corner of each page. TLFs should be internally paginated in relation to the total length (i.e., the page number should appear sequentially as page n of N, where N is the total number of pages in the table).
- The date (date output was generated) should appear along with program name and location as the last footer on each page.

12.2.3 Display Titles

Each TLF should be identified by the designation and a numeral. (e.g., Table 14.1.1). ICH E3 recommended numbering will be used. A decimal system (x.y and x.y.z) should be used to identify TLFs with related contents. The title is centered. The analysis set should be identified on the line immediately following the title. The title and table designation are single spaced. A solid line spanning the margins will separate the display titles from the column headers. There will be 1 blank line between the last title and the solid line.
Table x.y.z
First Line of Title
Second Line of Title (if needed)
Analysis Set

12.2.4 Column Headers

- Column headings should be displayed immediately below the solid line described above in initial upper-case characters.
- For numeric variables, include "unit" in column or row heading when appropriate.
- Analysis set sizes will be presented for each treatment group in the column heading as (N=xx) (or in the row headings if applicable). This is distinct from the 'n' used for the descriptive statistics representing the number of subjects in the analysis set.

12.2.5 Body of the Data Display

12.2.5.1 General Conventions

Data in columns of a table or listing should be formatted as follows:

- alphanumeric values are left-justified;
- whole numbers (e.g., counts) are right-justified; and
- numbers containing fractional portions are decimal aligned.

12.2.5.2 Table Conventions

- Units will be included where available
- Data will be presented in the following order (left to right or top to bottom, as appropriate) in tables: placebo (Cohort 2), orvepitant 10 mg, placebo (Cohort 1), orvepitant 30 mg.
- If the categories of a parameter are ordered, then all categories between the maximum and minimum category should be presented in the table, even if n=0 for all treatment groups in a given category that is between the minimum and maximum level for that parameter. For example, the frequency distribution for symptom severity would appear as:

Severity Rating	N
very severe	2
severe	0
moderate	8
mild	3

  Where percentages are presented in these tables, zero percentages will not be presented and so any counts of 0 will be presented as 0 and not as 0 (0%).
- If the categories are not ordered (e.g., Medical History, Reasons for Discontinuation from the Study, etc.), then only those categories for which there is at least 1 subject represented in 1 or more groups should be included.
- An Unknown or Missing category should be added to any parameter for which information is not available for 1 or more subjects.
- Unless otherwise specified, the estimated mean and median for a set of values should be presented to 1 more significant digit than the original values, and standard deviations should be presented to 2 more significant digits than the original values. The minimum and maximum should report the same significant digits as the original values. For example, for systolic blood pressure:

N	XX
Mean	XXX.X
Std Dev	X.XX
Median	XXX.X
Minimum	XXX
Maximum	XXX

- P-values should be output in the format: “0.xxx”, where xxx is the value rounded to 3 decimal places. Any p-value less than 0.001 will be presented as <0.001. If the p-value should be less than 0.0001 then present as <0.0001. If the p-value is returned as >0.999 then present as >0.999
- Percentage values should be printed to one decimal place, in parentheses with no spaces, one space after the count (e.g., 7 (12.8%), 13 (5.4%)). Values that round down to 0.0 will be displayed as ‘<0.1’. Unless otherwise noted, for all percentages, the number of subjects in the analysis set for the treatment group who have an observation will be the denominator. Percentages after zero counts should not be displayed and percentages equating to 100% should be presented as 100%.

- Unless otherwise specified, tabular display of data for medical history, prior / concomitant medications, and all tabular displays of adverse event data should be presented by the body system, drug class, or SOC in descending order, assuming all terms are coded. Within the body system, drug class and SOC, medical history (by preferred term), drugs (by ATC code), and adverse events (by preferred term) should be displayed in decreasing order. If incidence for more than 1 term is identical, they should then be sorted alphabetically. Missing descriptive statistics or p-values which cannot be estimated should be reported as "-".
- The percentage of subjects is normally calculated as a proportion of the number of subjects assessed in the relevant treatment group (or overall) for the analysis set presented. However, careful consideration is required in many instances due to the complicated nature of selecting the denominator, usually the appropriate number of subjects exposed. Details of this will be provided in footnotes in tables where appropriate.
- For categorical summaries (number and percentage of subjects) where a subject can be included in more than one category, a footnote will note if the subject should be included in the summary statistics for all relevant categories or just 1 category and the criteria for selecting the criteria.
- Where a category with a subheading (such as system organ class) has to be split over more than one page, output the subheading followed by "(cont)" at the top of each subsequent page. The overall summary statistics for the subheading will only be output on the first relevant page.

12.2.5.3 Listing Conventions

- Listings will be sorted for presentation in order of treatment groups (as above), subject number, visit/collection day, and visit/collection time.
- Missing data will be represented on subject listings as either a hyphen ("-") with a corresponding footnote ("- = unknown or not evaluated"), or as "N/A", with the footnote "N/A = not applicable", whichever is appropriate.
- Dates will be printed in SAS® DATE9.format ("DDMMMYYYY": 01JUL2000). Missing portions of dates will be represented on subject listings as dashes (--JUL2000). Dates that are missing because they are not applicable for the subject are output as "N/A", unless otherwise specified.
- All observed time values will be presented using a 24-hour clock HH:MM or HH:MM:SS format (e.g., 11:26:45, or 11:26). Time will only be reported if it was measured as part of the study.
- Units will be included where available

12.2.5.4 Figure Conventions

- Unless otherwise specified, for all figures, study visits will be displayed on the X-axis and endpoint (e.g., treatment mean change from baseline) values will be displayed on the Y-axis.

12.2.6 Footnotes

- A solid line spanning the margins will separate the body of the data display from the footnotes.

- All footnotes will be left justified with single-line spacing immediately below the solid line underneath the data display.
- Footnotes will always begin with "Note:" if an informational footnote, or 1, 2, 3, etc. if a reference footnote. Each new footnote will start on a new line where possible.
- Footnotes will be present on the page where they are first referenced and thereafter on each page of the table, unless the footnote is specific only to certain pages. Subject specific footnotes will be avoided wherever possible.
- Footnotes will be used sparingly and must add value to the table, figure, or data listing. If more than six lines of footnotes are planned, then a cover page may be used to display footnotes, and only those essential to comprehension of the data will be repeated on each page.
- The last 2 lines of the footnote section will be a standard source that indicates the name of the program used to produce the data display, date the program was run, and the listing source (or data source for a listing) (e.g., 'Program: myprogram.sas Listing source: 16.x.y.z').

13 QUALITY CONTROL

13.1 SPECIFICATIONS

Once the SAP is finalised, dataset and TLF specifications will be developed and reviewed by the study team. A round table review of draft specifications will be conducted by the [redacted] Lead Programmer, Lead Validator, Lead Statistician and Senior Reviewer (or member of Statistics Management).

NeRRe will then review, comment and approve (via signature) this analysis plan and all dataset and TLF specifications as well as any subsequent amendments to any of these.

13.2 OUTPUTS

Validation of analysis datasets and tables will be conducted through independent parallel programming of the statistical output according to the agreed upon specifications defined in the protocol, this SAP, table and shells, and dataset specifications. In this process, two programmers working independently (i.e., without input from one another), will program the same output and compare results (via SAS PROC COMPARE). Any discrepancies will be discussed and resolved, and the validation cycle repeated until no further differences are noted between the two outputs.

All programs will be submitted in batch mode to document the results of the PROC COMPARE indicating no unequal observations. Additionally, tracking logs will be maintained which document all QC and validation findings and their resolution.

For CDISC datasets the Pinnacle 21 report will be used to validate the datasets for CDISC compliance.

Once the validation cycle is complete, the output (dataset or TLF) will be subjected to the lead statisticians review as well as an [redacted] round table review including the Lead Programmer, Lead Validator, Lead Statistician and Senior Reviewer (or member of Statistics Management).

For the delivery of the outputs for the Blinded Data Review Meeting (BDRM) the following team members will join the output review round table as well: Data Manager, Medical Monitor and Project Manager.

NeRRe will review, comment and approve (via signature) all final datasets and TLFs.

14 APPENDICES

14.1 CHANGES TO THE PROTOCOL SPECIFIED ANALYSES

None.

14.2 SAMPLE SIZE RE-ESTIMATION PLAN

After 24 subjects have completed Treatment Period B (approximately 12 subjects per cohort), the variance of the cough severity NRS data will be reviewed on a fully blinded basis. Up to a maximum of 20 subjects (approximately 10 per cohort) may be recruited if the emerging cough severity NRS data is found to have greater variance than allowed for in the original sample size estimate. Any possible impact on the type I error due to this blinded sample size re-assessment procedure is believed to be negligible.

Following Graylin (2018), the null adjusted blinded estimator of the within-subject variance will be used to re-estimate sample size. The Cohort and treatment sequence assignment within cohort is blinded. The within-subject variance is assumed to be the same for both cohorts.

PROC UNIVARIATE will be used to compute estimates of the standard deviation of the paired difference between treatment periods (period 2 minus period 1). Estimates from PROC UNIVARIATE will include the ordinary point estimate of the standard deviation and various estimates derived from robust estimates of scale (obtained by using the ROBUSTSCALE option in PROC UNIVARIATE) which are less influenced by outliers. See Table Template 1 for a list of the robust estimates of scale. If there is consistency between the ordinary point estimate and the robust estimators, then this supports the reliability of the ordinary point estimate of the standard deviation. If the estimates are not consistent, then this indicates the influence of outliers and/or departures from normality on the estimates and thus results should be interpreted with caution. In both cases, the estimates provide a range of plausible values for the standard deviation. An 80% confidence interval of the point estimate of the standard deviation is also provided to account for uncertainty in the ordinary point estimate and to provide a highly plausible range of values for the standard deviation.

The following SAS code will be used to compute sample size estimates for the various estimates of the standard deviation of the paired difference, as described above.

```
proc power;
   pairedmeans test=diff
      meandiff = 1.5
      corr = 0.5
      stddev = <standard deviation of the paired difference>
      npairs = .
      power = 0.9;
run;
```

Programming notes for the power calculation:

The within-subject variance is equal to the variance of the paired difference between treatment periods divided by 2.

To use PROC POWER for sample size calculations, we assume a common standard deviation, SD, for the primary endpoint in each treatment period. If RHO represents the correlation of the primary endpoint between the two periods, then SD is equal to the within-subject standard deviation divided by sqrt(1-RHO). Sample size is invariant to the selection of RHO as long as SD is derived as the within-subject standard deviation divided by sqrt(1-

RHO). Without loss of generality, RHO may be set to 0.5 so that SD is equal to sqrt(2) * within-subject standard deviation or equivalently SD is equal to the standard deviation of the paired difference.

Outputs similar to Table Template 1 and Figure Template 1 will be produced for the BSSR. No other tables, figures, or listings will be produced for the BSSR.

- Table Template 1 will be used to display sample size estimates based on various estimates of the standard deviation of the paired difference. This includes the protocol-specified assumption for the standard deviation (i.e. 2.7) and various estimates of the standard deviation based on the data cut.
- Figure Template 1 will be produced to assess the distribution of paired differences.

The BSSR outputs will be based on evaluable subjects, defined as subjects who were randomized and received treatment and have sufficient data to compute the primary endpoint for both treatment periods. The primary endpoint for both periods can be computed only if all the following conditions are met:

- The subject has 4 study visits within the windows set forth in Section 5.3.2 of the SAP for analysis visits corresponding to Visit 2, Visit 4, Visit 5, and Visit 7. This condition is met when the study day for a visit is within the protocol window (Column A) or within the "reasonable limits for out of window days" (Column B).
- The subject has at least 4 daily IPF cough severity scale scores in the 7 calendar days up to and including the date of each relevant visit (i.e. the analysis visits for Visit 2, Visit 4, Visit 5, and Visit 7).

Review

If the estimates of standard deviation, as summarised in the output similar to Table Template 1, show a range of possible outcomes with respect to sample size, the data will be reviewed and a consensus decision reached on the new sample size to be applied to the study.

References

Grayling MJ, Mander AP, Wason JMS. Blinded and unblinded sample size reestimation in crossover trials balanced for period. Biom J. 2018;60:917–33.

Hampel, F. R. (1974). "The Influence Curve and Its Role in Robust Estimation." Journal of the American Statistical Association

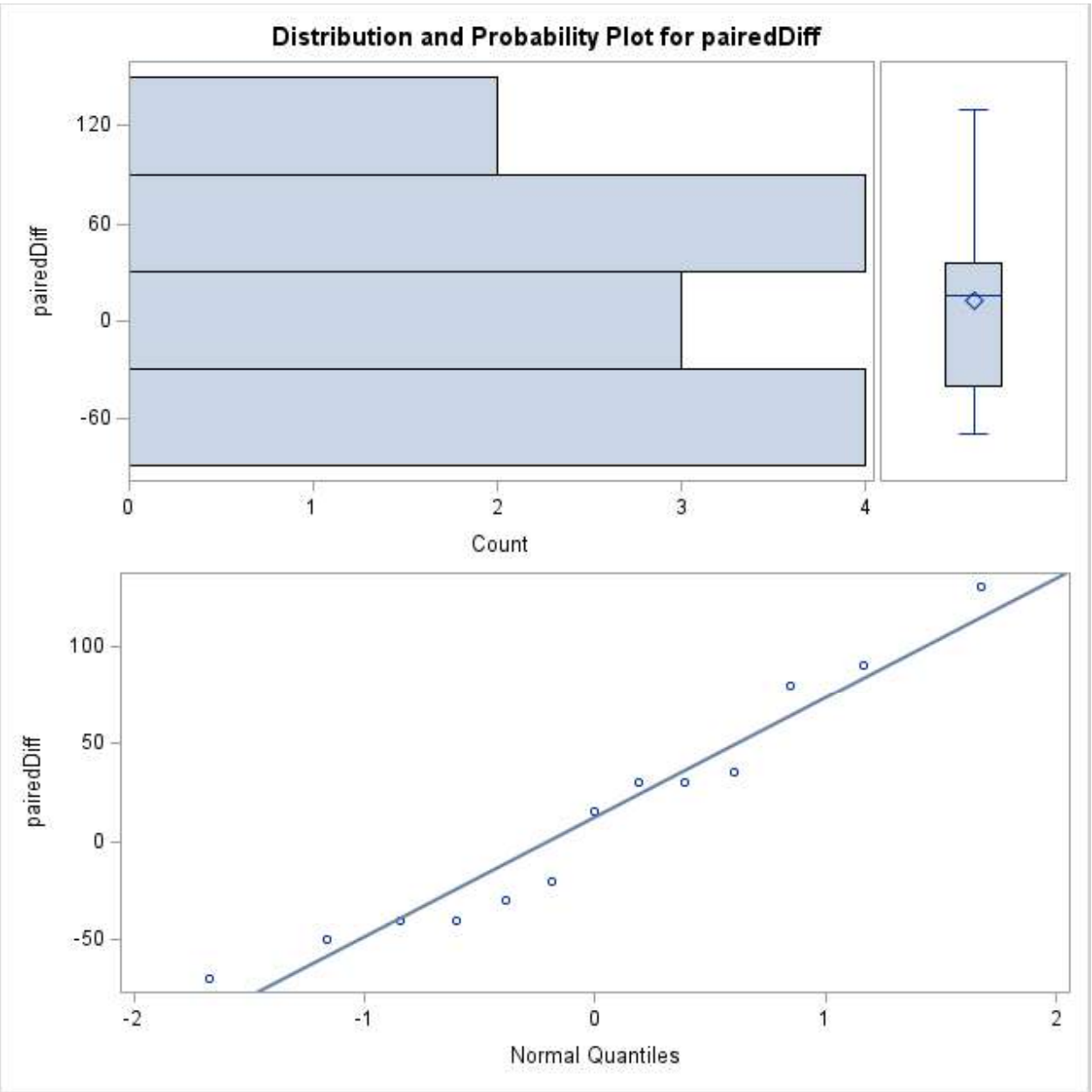
Croux, C., and Rousseeuw, P. J. (1992). "Time-Efficient Algorithms for Two Highly Robust Estimators of Scale." Computational Statistics

Table Template 1: Blinded Sample Size Re-estimation [1]
Number of Subjects in Both Cohorts with Evaluable Data [2] (N=XX)
Data as of: dd-mmm-yyyy

Type of Estimate	Standard Deviation of Paired Difference	Sample Size Estimate per Cohort
Protocol-Specified	2.7	XX
Point Estimate	XX.XX	XX
80% Confidence Interval of Point Estimate [3]		
Lower Bound	XX.XX	XX
Upper Bound	XX.XX	XX
Estimates derived from Robust Measures of Scale [4]		
Interquartile Range / 1.34898	XX.XX	XX
Gini's Mean Difference * sqrt(pi)/2	XX.XX	XX
MAD * 1.4826	XX.XX	XX
Sn * Csn	XX.XX	XX
Qn * Cqn	XX.XX	XX

Pi is the numerical constant that represents the ratio of a circle's circumference to its diameter.
MAD is Median absolute deviation about the median (Hampel; 1974).
Sn and Qn are robust estimators proposed by Croux 1992. Csn and Cqn are correction factors used to estimate the standard deviation from Sn and Qn, respectively.
[1] Sample size estimates per cohort assume a treatment difference=1.5, two-sided alpha=0.05, and 90% power using a paired t-test.
[2] A subject is evaluable if they were randomized and received treatment and the primary endpoint can be computed for both treatment periods.
[3] This confidence interval assumes (n-1)*Vhat/V is chi-square with n-1 df, where Vhat and V are the variance estimate and the variance parameter, respectively.
[4] See SAS documentation for the ROBUSTSCALE option of PROC UNIVARIATE.

Figure Template 1: Distribution and Probability Plot of the Blinded Paired Differences
Number of Subjects in Both Cohorts with Evaluable Data [2] (N=XX)
Data as of: dd-mmm-yyyy



[1] A subject is evaluable if they were randomized and received treatment and the primary endpoint can be computed for both treatment periods.