Assessment of the effect of Positive Airway Pressure on energy and vitality in mild Obstructive Sleep Apnea patients

The Merge Study

Statistical Analysis Plan

Version 1.5 – 25Sep2019
Based on version 4.0 of protocol (25 May 2018)

ClinicalTrials.gov NCT: 02699463

Table of Contents

1. INTRODUCTION

This document details the proposed presentation and analysis for the main paper(s) reporting results from the ResMed funded multi-centre, unblinded, randomised, controlled trial of the effect of positive airway pressure on energy and vitality in mild Obstructive Sleep Apnoea (OSA) patients (The MERGE study). The results reported in these papers should follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial.

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this statistical analysis plan.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.1 Trial Statisticians

Leslee Willes, MS
Willes Consulting Group, Inc.
Encinitas, California USA 92024
lesleew@willesconsulting.com

Colleen Kelly, PhD
Kelly Statistical Consulting, Inc.
Carlsbad, California USA 92011
kstat.consulting@gmail.com

2. CHANGES FROM PREVIOUS VERSION OF SAP

The Statistical Analysis Plan (SAP) has been re-evaluated and modified since Version 1.2, because of a change in statistical representation for study oversight.

3. GLOSSARY OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
AASM	American Association of Sleep Medicine
BMI	Body Mass Index
CC	Completed Cases

ABBREVIATION	DESCRIPTION
CPAP	Continuous Positive Airway Pressure
Euroqol EQ-5D	5 Dimension Health Questionnaire
ESS	Epworth Sleepiness Scale
FOSQ	Functional Outcomes of Sleep Questionnaire
FSS	Fatigue Severity Scale
HADS	Hospital Anxiety and Depression Scale
HST	Home Sleep Test
ISI	Insomnia Severity Index
ITT	Intention to Treat
MAR	Missing at random
MCS	Mental Component Summary
ORTU	Oxford Respiratory Trials Unit
OSA	Obstructive Sleep Apnoea
PCS	Physical Component Summary
PP	Per Protocol
PROM	Patient Reported Outcome Measure
SAP	Statistical Analysis Plan
SF-36	Short Form (SF) – 36 Questionnaire
TSC	Trial Steering Committee
VAS	Visual Analog Scale

4. STUDY DESIGN

4.1 Background

Patients with Obstructive Sleep Apnoea (OSA) experience interrupted and reduced sleep which can have a significant impact on quality of life and increases the risk of developing certain conditions. Mild OSA, unlike moderate and severe OSA, has not been extensively studied. Although there is a reasonable pool of evidence that suggests that even minor sleep related breathing disturbances are associated with negative consequences, there is no consensus on when treatment should be initiated or the best treatment for mild OSA patients. Two previous clinical trials have suggested that treating mild OSA with Continuous Positive Airway Pressure (CPAP) improved symptoms, quality of life and reduced health risks in these patients. However, despite this evidence, treatment with CPAP for mild OSA is not consistently offered.

CPAP is a treatment that uses mild air pressure to keep the airways open during sleep. The machine (ResMed AirSense S10) consists of a mask that fits over the nose, a tube that connects the mask to the machine motor and the motor which blows air through the tube to the mask.

Mild OSA is defined as Apnoea Hypopnoea Index (AHI) of 5 up to 15 events/hour. Mild OSA patients will be identified using **American Academy of Sleep Medicine (**AASM) 2007 and 2012 guidelines as described below.

In 2012, the American Academy of Sleep Medicine (AASM) changed the criteria needed to score a hypopnoea during a sleep study. Previous 2007 criteria had required a hypopnea to include a decrease in oronasal airflow by ≥ 30% from baseline, an event of at least 10 seconds, and ≥4% SpO2 desaturation. These criteria had been widely accepted by healthcare providers to form the basis of

diagnosis, access to treatment, and reimbursement for patient equipment. In 2012, a change was made to these criteria which permitted hypopnoeas to be scored with an arousal only, or an oxygen desaturation of ≥3%. The newer criteria of OSA defined by the AASM in 2012 are broader and increase the percentage of patients who may be diagnosed with OSA, potentially by up to 40%.

Guideline	Recommended Hypopnea Definition per Guidelines:
AASM 2007	• A decrease in oronasal airflow by ≥ 30% from baseline AND • The event is ≥ 10 sec long AND • Associated with ≥ 4% SpO_2 desaturation
AASM 2012	• A decrease in oronasal airflow by ≥ 30% from baseline AND • The event is ≥ 10 sec long AND • Associated with ≥ 3% SpO_2 desaturation OR arousal

However, the new scoring rules remain controversial as there is no compelling evidence that patients with mild OSA according to this new criterion benefit from treatment. As a result, health insurance reimbursement rules in some countries have not adopted the 2012 criteria.

In the MERGE Trial, the patients are initially screened for eligibility based on automatic AASM 2007 scoring. Those who meet the criteria for Mild OSA (AHI 5-15 events/hr) based on AASM 2007 scoring using the ApnoeaLink home sleep test (automatically scored by Airview software) are eligible for enrolment. Those patients who have no OSA (AHI < 5 events/hr) per AASM 2007 scoring are rescored using AASM 2012 automated scoring. Those patients who meet the criteria of AHI ≥ 5 events/hr per AASM 2012 are also eligible for enrolment.

4.2 Enrolment Target and Inclusion/Exclusion Criteria

The MERGE study is a prospective multicentre randomised parallel trial with 3 months of active treatment versus a control arm. Randomisation with minimisation will be performed with a 1:1 treatment group allocation.

Date of start of recruitment:	26 October 2016
Date of expected end of recruitment:	31 December 2018
Date expected end follow-up:	31 March 2019
Target number of subjects:	224 (112 per arm)
Participating Centres:	11

A patient will be eligible for inclusion in the study if all the following inclusion criteria apply:

- AHI 5-15 events/hr as per AASM 2007 scoring criteria or AHI ≥5 per AASM 2012 if AHI 0-4.9 per AASM 2007
- Aged ≥ 18 and ≤ 80 years
- Ability and willingness to provide written informed consent
- Ability to tolerate a CPAP one hour long run-in test

A patient will not be eligible for the study if any of the following exclusion criteria apply:

- The presence of unstable cardiac disease at trial screening
- Inability to give fully informed consent
- Use of supplemental oxygen
- Secondary sleep pathology e.g. Periodic Limb Movement Syndrome, Narcolepsy, Circadian Disorder, Obesity Hypoventilation Syndrome
- Epworth Sleepiness Scale (ESS) ≥ 15, or concerns about sleepy driving from physician/ sleep lab staff
- Body Mass Index (BMI) ≥ 40 kg/m^2
- Previous CPAP usage

4.3 Treatment Groups

The Treatment group will receive CPAP along with standard care for mild OSA patients (counselling on healthy lifestyle behaviours and sleep hygiene habits). The Control group will receive only standard of care for OSA patients.

4.4 Description of Adherence in Treatment Group

Adherence with CPAP in the Treatment group will be defined as follows. Patients in the Treatment group will be considered adherent to CPAP if the patient uses the device at least 4 hours per day on average during the 3-month study period.

The US Medicare definition for adherence will be considered as a secondary definition for exploratory analyses. By this definition, patients who use the device at least 4 hours per night for 70% of the nights during a 30-day period over the 3-month study period are considered adherent.

Adherence to CPAP will be monitored using ResMed AirView Remote Monitoring system. Daily usage data over the study period will be provided for analysis.

4.5 Study Objectives

This study aims to assess the changes in quality of life from baseline to 3 months in mild OSA patients (based on AASM 2012 guidelines) comparing the Treatment and Control groups.

The objectives of the study are as follows:

Primary Objective:

- Compare the change in quality of life (SF-36 energy and vitality subscale score) from baseline (pre-treatment) to 3 months between Treatment and Control groups in patients with mild OSA **per AASM 2012** guidelines.

Secondary Objectives:

- Compare changes in quality of life scores (SF-36, ESS, FSS, FOSQ, HADS, ISI and EQ-5D) from baseline (pre-treatment) to 3 months between Treatment and Control groups in patients with mild OSA **per AASM 2012** guidelines.
- Compare changes in quality of life scores (SF-36, ESS, FSS, FOSQ, HADS, ISI and EQ-5D) from

baseline (pre-treatment) to 3 months between Treatment and Control groups in patients with mild OSA **per AASM 2007** guidelines.

Exploratory Analyses:

- Compare changes in quality of life scores (SF-36, ESS, FSS, FOSQ, HADS, ISI and EQ-5D) from baseline (pre-treatment) to 3 months between Treatment and Control groups in patients with moderate to severe OSA **per AASM 2012** guidelines**.**

- Compare changes in quality of life scores from baseline to 3 months between Treatment and Control groups for patients with no OSA (AHI<5) per **AASM 2007 guidelines** but who enter the study (AHI≥5) per **AASM 2012 guidelines.**

- Assess the impact of adherence to CPAP in the Treatment group, comparing changes in quality of life scores from baseline to 3 months for patients adherent and not adherent to CPAP over the course of study.

- Explore variation in the treatment effect on the change in quality of life scores (SF-36 energy and vitality subscale score, ESS, and FSS) from Baseline to 3 months over the continuous baseline measures of age, BMI, and AHI (as a continuous measure of OSA severity), separately for AASM 2007 and AASM 2012 guidelines.

- Assess the number of patient telephone contacts, clinic visits and other types of intervention required to support CPAP adherence, including the impact of missed contacts, in the Treatment group, comparing results between patients adherent and not adherent to CPAP.

- Compare baseline symptoms and changes in quality of life scores from baseline to 3 months between Treatment and Control groups by gender.

- Compare change from baseline to 3 months in weight and BMI between Treatment and Control groups**.**

- A cost effectiveness and health economic analysis will be performed by an outside consultant. Details of this analysis will be provided in a separate analysis plan.

4.6 Sample Size

The sample size of 224 (112 per arm) is based on statistics observed in the MOSAIC study: a mean score difference of 6.6 was observed between 3 months and baseline (considered a significant clinical improvement by the authors of the study) in the Energy and Vitality subscale of the SF-36 questionnaire, with a mean change of 4.2 (SD 18.1) in the Control group and 10.8 (SD 17.0) in the CPAP Treatment group. The sample size calculation is based on a two-sided 5% significance level and 80% power.

It is estimated that approximately 300 participants will need to be randomised in order to reach the required sample size of 224 patients with mild OSA per AASM 2012 guidelines. Subjects who are eligible for Home Sleep Testing (HST) are evaluated for their level of OSA. Those patients who have Mild OSA per AASM 2007 guidelines are eligible for randomisation. Those patients who have No OSA per AASM 2007 guidelines are further assessed based on the AASM 2012 guidelines. All patients with

AHI≥5 per AASM 2012 guidelines are also eligible for randomisation. Since the primary endpoint analysis is based on patients with mild OSA per AASM 2012 guidelines, there will be some patients randomised who will not fall into this sample, thus the need for additional enrolment. Additionally, it is expected that 10% of those randomised will drop-out before the end of the study.

The number of patients required to enrol will be reviewed and updated by the Trial Steering Committee (TSC) as new recruitment information accumulates. This will not change the goal of 224 patients completing the study who had mild OSA per the AASM 2012 guidelines.

4.7 Randomisation

Randomisation was carried out using the RRAMP randomisation system provided by the Oxford Respiratory Trials Unit (ORTU). Minimisation method was used to allocate patients to treatment and control groups, maintaining balance in the following three stratifying factors:

- Gender: male; female.
- Age: <30 years, 30-60 years, >60 years
- BMI: <30 kg/m^2, ≥30 kg/m^2

4.8 Definitions of Study Endpoints

Patient outcome measures are generated from six quality of life questionnaires. All outcome measures are Patient Reported Outcome Measures (PROMs). Each questionnaire, when fully completed provides one or more summary scores as detailed below.

1. **Short Form 36 (SF-36):** The Short Form (SF-36) Health Survey is a 36-item, patient-reported survey of patient health consisting of eight subscales and two summary scores. The survey determines the general quality of life of a patient based on 8 subscales (Bodily Pain, Energy/Vitality, General Health, Mental Health, Physical Functioning, Role Emotional, Role Physical and Social Functioning), and a physical component summary (PCS) score and mental component summary (MCS) score. Secondary analysis will use both the 8 subscales and the MCS and PCS summary scores. These scores will be calculated using PRO CoRE Smart Measurement® System software developed by OPTUM® and presented as norm-based scores.
2. **Epworth Sleepiness Scale (ESS):** The ESS form is a measure of a patient's likelihood to fall asleep in eight different scenarios. It is used in sleep clinics to detect dangerous levels of sleepiness in patients with sleep disorders and the summary score is an integer value of the total score of all questions ranging from 0 to 24.
3. **Fatigue Severity Scale (FSS):** The FSS is designed to measure the severity of fatigue and how it impacts an individual's daily life in terms of motivation, exercise, daily activity, work, family and social life. It consists of 9 questions with a total score ranging from 9 to 63, with higher scores associated with more fatigue. Secondary analysis will be performed on the total FSS score.
4. **Functional Outcomes of Sleep Questionnaire (FOSQ):** The FOSQ is a questionnaire designed to assess the impact of excessive sleepiness on multiple activities of everyday living. The FOSQ contains 30 items summarized as one total score and five subscales (Activity Level, Vigilance, Intimacy Relationships and Sexual Activity, General Productivity, and Social Outcome). Secondary analysis will be performed on the total FOSQ score.
5. **Hospital Anxiety and Depression Scale (HADS):** The HADS questionnaire contains 14 items in

which patients must respond on a scale of 0 to 4 to describe their current state of being. Two scores are calculated, one for depression and one for anxiety, both ranging from 0 to 21 with larger scores representing higher levels of anxiety/depression. Secondary analysis will be performed separately on the Anxiety and Depression scores.

6. **Insomnia Severity Index (ISI):** The ISI seven-item questionnaire asks the individual to rate the level of insomnia and perceptions of sleep problems in the last two weeks on a scale of 0 to 4. The total score ranges from 0 to 28 with larger scores indicating more severe insomnia. Secondary analysis will be performed on the total ISI score.
7. **European Quality of Life 5 Dimensions Questionnaire (Euroqol EQ-5D)**: The Euroqol EQ-5D questionnaire consists of five dimensions assessing mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 response levels ranging from no problems to extreme problems. The five dimensions are compiled to generate an EQ-5D Index. The second part of the questionnaire uses a Visual Analog Scale (VAS) to assess the patient's current health status on a scale of 0 to 100. Secondary analysis will be performed on the EQ-5D Index and VAS score.

4.9 Primary Hypothesis and Study Endpoints

Primary Study Hypothesis

The CPAP Treatment group will have greater improvement on the SF-36 Energy and Vitality subscale score in patients with mild OSA **per AASM 2012 guidelines** compared to the Control group.

Primary Endpoint

Change in the Energy and Vitality Subscale Score of the SF 36 questionnaire from baseline (pre-treatment) to 3 months in mild OSA patients **per AASM 2012 guidelines**.

Secondary Endpoints

Change in outcome scores between baseline (pre-treatment) and 3 months in mild OSA patients **per AASM 2012 guidelines** in using the following quality of life measures:

SF-36 (7 subscales not previously assessed as the primary endpoint, PCS and MCS summary scores)
ESS
FSS total score
FOSQ total score
HADS Anxiety and Depression scores
ISI total score
Euroqol EQ-5D index
VAS score

Change in outcome scores between baseline (pre-treatment) and 3 months in mild OSA patients **per AASM 2007 guidelines** in the following quality of life measures:

SF-36 (8 subscales, PCS and MCS summary scores)
ESS

FSS total score
FOSQ total score
HADS Anxiety and Depression scores
ISI total score
Euroqol EQ-5D index
VAS score

4.10 Outcomes Assessment Schedule

The following outcome measures will be assessed at baseline (before randomisation) and 3 months post treatment initiation.

Outcome	Baseline	3 days	3 months
Demographics & History	✓		
Home Sleep Test	✓		
Troubleshoot Phone Call		✓	
SF-36	✓		✓
ESS	✓		✓
FSS	✓		✓
FOSQ	✓		✓
HADS	✓		✓
ISI	✓		✓
Euroqol EQ-5D	✓		✓
CPAP Questionnaire			✓

5. STATISTICAL METHODS

5.1 Study Populations and Analysis Groups

There are two main study populations in this trial, one group fulfilling the mild OSA AASM 2012 guidelines and the second fulfilling the mild OSA AASM 2007 guidelines. Some trial patients will fulfil both criteria and will therefore be eligible for analysis under both AASM 2007 and AASM 2012. A third study population will consist of patients who fall into the category of Moderate to Severe OSA based on the AASM 2012 guidelines. Patients in this third population will either have tested as No OSA (AHI < 5 e/hr) or Mild OSA based on AASM 2007 guidelines.

The primary endpoint analysis will only be conducted on mild OSA patients per AASM 2012 guidelines. Each secondary endpoint analysis will be carried out on mild OSA patients based on AASM 2012 and AASM 2007 guidelines. Additionally, secondary analyses will be generated descriptively for the subgroup of patients with Moderate to Severe OSA per AASM 2012 guidelines. Therefore, all patients randomised into the trial will be analysed in one or more secondary analyses, with a subset of these patients fulfilling the AASM 2012 guidelines and qualifying for the primary analysis.

The intention to treat (ITT) analysis population will include all randomised patients. The ITT population will be further subdivided as follows:
ITT-1 population includes patients in the ITT with mild OSA per AASM 2012 guidelines.

ITT-2 population includes patients in the ITT with mild OSA per AASM 2007 guidelines.
ITT-3 population includes patients in the ITT with moderate to severe OSA per AASM 2012 guidelines.

The primary endpoint analysis, pool-ability analysis (see Section 5.3) and comparability of baseline characteristics will be generated for the ITT-1 population.

The completed cases (CC) analysis population will be a subset of the ITT population that includes patients who complete the 3-month study visit. [It is possible for there to be missing values for one or more outcome measures in the CC group, if a patient who completes the visits does not complete the questionnaires completely.] The CC population will be further subdivided as follows:
CC-1 population includes patients in the CC with mild OSA per AASM 2012 guidelines.
CC-2 population includes patients in the CC with mild OSA per AASM 2007 guidelines.
CC-3 population includes patients in the CC with moderate to severe OSA per AASM 2012 guidelines.

The secondary endpoint analyses will be generated for the CC-1, CC-2 and CC-3 populations, as specified.

5.2 Multiplicity

There is only one primary hypothesis in this study and therefore, no adjustments are required for multiple primary hypotheses.

5.3 Data Pooling

The issue of pool-ability of the data entails two features: the combining of data across study sites and the method of computation of an overall estimate for the primary study endpoint. The justification for combining of data across sites is based on the clinical assessment provided by Meinert (1986): the clinical study will be conducted under a common protocol for each investigational site, the study sites will be monitored for protocol compliance, and the same data gathering instruments and methods will be used in every site.

In order to test homogeneity of the primary endpoint across study sites and account for possible differences in demographic characteristics at study sites, a mixed-effects regression model for the ITT-1 population will be constructed with the change in the Energy and Vitality subscale of the SF-36 questionnaire from baseline (i.e., pre-treatment) to 3 months post-randomization as the dependent variable and will include site, treatment group and a site-by-treatment group interaction as a fixed effects. If the treatment-by-site interaction is statistically significant, this may imply that the treatment effect differs across study sites. The treatment-by-site interaction will be tested at a significance level of 0.10. If a treatment-by-site interaction is significant, a mixed-effects regression model with the following additional covariates will be considered to see if they help to explain the treatment-by-site interaction: baseline Energy and Vitality score, percentage adherent for the Treatment group at each site, and other baseline and demographic variables that are unbalanced across study sites.

In the case that there is a statistically significant treatment-by-site interaction, meaning that the treatment effect is heterogenous across study sites, then study site and treatment*site will be

included as random effects in the mixed-effects repeated measures model for the primary endpoint analysis.

The pool-ability analysis of study sites may require the formation of pseudo-sites because small sites will not provide appropriate information to allow the analysis above. The formation of pseudo-sites will be done without regard to effect size. Sites will be combined into pseudo-sites in an unbiased way, using the sorted study site numbers. Study sites with fewer than 6 patients will be combined in the following manner. The first site with less than 6 patients will be combined with the next site numbered in consecutive order with less than 6 patients. If this combination produces a size for the combined sites that is lower than the mean size for all study sites in the study, the next site will be added, as long as the total is less than or equal to the mean size for the study. If that site makes the size bigger than the mean size, the next site will be considered for inclusion in the pseudo-site. If no additional sites can be found that can be added and the size remain less than or equal to the mean of the study site sizes, that pseudo-site will be closed and another will be started with the first site that could not be added to the previous pseudo-site. This process continues until all small sites are combined into pseudo-sites. The remaining study sites with size of 6 or more patients will be assigned pseudo-site numbers. If there is but one site remaining with less than 6 patients, that site will be added to the smallest of the study sites that have at least 6 patients.

5.4 Missing Data

The number and percentage of patients with missing data for each outcome measure at each time point will be reported by treatment group. A completed outcome measure refers to a completed questionnaire with a non-missing summary score. Questionnaires where too many questions were left blank, hindering the calculation of one or more summary scores, will be counted as a missing outcome measure for the corresponding endpoint.

The primary analysis uses a mixed-effects repeated measures model, which assumes data is missing at random (MAR) and incorporates all available information from all subjects.

Baseline characteristics will be summarised for those who did and did not complete the SF-36 energy and vitality questions at the 3-month follow-up visit, by treatment group, to describe any characteristics related to missingness that can be observed.

5.5 Patient Accountability

A CONSORT flowchart will be constructed to summarise the progression of patients through each study stage for each study population, and reasons for drop-out or exclusions at each stage. The flowchart will show the number of patients in each treatment group who (1) met the initial study criteria, (2) were randomly assigned, (3) received intended treatment, (4) completed specified visits, and (5) were analysed for the primary endpoint. The flowchart will present results overall and for each study population (mild OSA per AASM 2012, mild OSA per AASM 2007 and moderate/severe OSA per AASM 2012).

Protocol deviations from will be presented descriptively by treatment group with the number and percent of total deviations for each deviation type. Protocol deviations include:
Missed 3-month visit
Out of window 3-month visit (visit window defined as 90 days + 1 week)
Incomplete or missing SF-36 Energy and Vitality question responses

The numbers (with percentages) of patients completing the study and lost to follow-up (defaulters and withdrawals) over the 3-month period of the study will be presented by treatment group. Additionally, reasons for losses will be summarized.

5.6 Comparability Analysis of Baseline Characteristics

Patients in the two treatment groups will be compared with respect to baseline and demographic characteristics (e.g., ethnicity, medical history), minimisation factors for randomisation (i.e., gender, age, BMI), and baseline quality of life endpoints for the ITT-1 population.

Baseline sleep test parameters will be compared by treatment group for the ITT-1 and ITT-2 populations.

Descriptive statistics within each treatment group will be presented for demographic and baseline values. For binary and categorical variables, the number evaluated, counts (with percentages) and 95% exact confidence limits for the percentage will be presented. For continuous variables, the number evaluated, mean, standard deviation, median, minimum, maximum, and 95% confidence interval for the mean will be presented. Treatment groups will be compared with respect to baseline variables in order to test that the groups are balanced. For continuous variables, a student's t-test or Wilcoxon rank sum test will be used, as appropriate. For categorical and binary endpoints, Fisher's exact test will be utilized. All tests will be tested using a Type I error rate of 0.05.

5.7 Primary Endpoint Analysis

The primary endpoint measure is the change in the Energy and Vitality subscale of the SF-36 questionnaire from baseline (i.e., pre-treatment) to 3 months post-randomization. It is a continuous variable.

The primary endpoint will be analysed using a mixed-effects repeated measures model to account for missing values. The model will be set up as follows: The Energy and Vitality subscale is the dependent variable, visit (baseline and month 3) is the repeated measure (a fixed effect), subject and site are random effects (with subject nested in site) and treatment group by visit is a fixed effect. In the case that the primary endpoint was not homogeneous across study sites (as discussed in Section 5.3), a treatment by visit by site interaction will also be included in the primary endpoint model as a random effect. Results will be presented as the mean difference in the change score from baseline to 3 months (adjusted for baseline score, as appropriate) between the two treatment groups with a 95% confidence interval at 3 months and p-value for the treatment group effect.

As a sensitivity analysis of the primary endpoint, a similar mixed-effects repeated measures model to the above will be constructed, that also adjusts for other baseline and demographic variables that were found to be imbalanced between treatment groups per Section 5.6. The mean difference in change score from baseline to 3 months will be presented with a 95% confidence interval at 3 months.

For these analyses of the primary endpoint, a P-value of less than 0.05 will be used to indicate statistical significance. Analyses of the primary endpoint will be generated for the ITT-1 population.

A second sensitivity analysis for the primary endpoint will be generated using the CC-1 population.

5.8 Secondary Endpoint Analyses

A mixed-effects repeated measures model will be utilized to compare treatment groups for all

secondary endpoints to account for missing values, as needed. The model will be set up as follows: The Secondary Endpoint subscale or total score is the dependent variable, visit (baseline and month 3) is the repeated measure (a fixed effect), subject and site are random effects (with subject nested in site) and treatment group by visit is a fixed effect.

Each secondary endpoint measure will be analysed in the same way as the primary endpoints for comparing the Treatment and Control groups using the CC-1 and CC-2 populations. Results will be presented as the mean difference in the change score (for each secondary endpoint) from baseline to 3 months (adjusted for baseline scores, as appropriate) between the two treatment groups with a 95% confidence interval at 3 months and p-value for the treatment group effect.

All secondary endpoints will be analysed separately for patients in the CC-1 and CC-2 populations There will be no statistical testing of differences between the CC-1 and CC-2 populations.

There will be no formal adjustments for multiple significance testing. All secondary analyses will be considered exploratory. We will explicitly state that the study was not powered for statistical comparisons of secondary endpoints in study publications and will urge caution for the interpretation of results obtained using secondary endpoints.

5.9 Safety Analyses

Any adverse event occurring while a patient is continuing in the study through the 3-month visit will be recorded. A comparison of adverse events between the Treatment and Control groups will be assessed descriptively, presenting the number of events and percent of patients having one or more of each event type by treatment group. A listing of all adverse events reported during the study will be presented, including the event description, time of occurrence, severity, outcome, expectedness and relationship to the study treatment. The safety analysis will be conducted on the ITT population.

5.10 Additional Exploratory Analyses

All secondary endpoints will be summarized by treatment group and presented descriptively for the CC-3 population (CC patients diagnosed as Moderate to Severe OSA per AASM 2012 guidelines). Descriptive statistics for the mean change from baseline to the 3-month visit for each endpoint include sample size, mean, standard deviation, median, minimum and maximum values for each endpoint.

Changes in quality of life scores from baseline to 3 months will be compared between Treatment and Control groups for patients with no OSA (AHI<5) per **AASM 2007 guidelines** but who enter the study (AHI≥5) per **AASM 2012 guidelines.** Descriptive statistics will be presented for the CC-1 population, and include sample size, mean, standard deviation, median, minimum and maximum values for each endpoint.

All secondary endpoints will be evaluated for the CPAP Treatment group, comparing patients in the Treatment group who are adherent to CPAP and not adherent to CPAP using descriptive statistics. No statistical inference will be made comparing these two groups and the results will be presented separately for patients in the CC-1 and CC-2 populations. Separate analyses will be presented for each definition of CPAP adherence.

Adherence to CPAP in the Treatment group will be further explored using logistic regression to assess the impact of baseline characteristics, OSA severity and other potential factors that predict compliance.

Usage data downloaded from the CPAP machines will also be summarized for all patients in the Treatment group over the 3- month study period, including residual AHI, mean usage hours per day, % of days with more than 4 hours of usage, median 95th percentile mask leak and % of patients using the myAir application. Additionally, the number of phone calls, emails, patient visits, reasons for contacts and missed contacts will be summarized to assess the effort required to support CPAP adherence. Comparisons will be presented overall and by CPAP adherence for patients in the ITT-1 and ITT-2 Treatment group populations.

The primary endpoint measure (mean change in Energy and Vitality subscale) will be examined over the spectrum of baseline AHI scoring as a continuous variable to investigate trends in the relationship between the two variables. Scatter plots of baseline AHI versus the primary endpoint will be generated for each treatment group. Statistical modelling may be performed to describe any observed variation in treatment effect related to baseline AHI. The primary endpoint will also be examined for any relationship between treatment effect and age or baseline BMI, using the methods described for AHI. These summaries will be produced for the ITT-1 population.

Mean change in ESS and FSS will also be assessed for variation in treatment effect related to baseline AHI, age, or BMI, using the methods described for the primary endpoint.

The primary endpoint measure will also be compared over the spectrum of baseline ODI and RERA scoring in a similar fashion as above. Separate scatter plots will be generated for the Treatment group, distinguishing patients adherent and non-adherent to CPAP.

Baseline and demographic characteristics, AHI, and quality of life scores will be summarized descriptively by gender for the ITT-1 population. Changes in quality of life scores from baseline to 3 months will be summarized descriptively by treatment group and gender, and the effect of gender will be explored by including gender as a fixed effect in the analysis model described in Section 5.7.

Change from baseline to 3 months in weight and BMI will be summarized descriptively by treatment group for the ITT-1 population.

The reason for patient referral will be analysed, comparing the primary endpoint measure by reason for referral separately for each treatment group using descriptive statistics.

5.11 Changes to Planned Analyses

Automatic sleep test scoring and manual review of scoring results will not be compared as part of this study analysis. This outcome is the primary focus of the Rich Berry Trial, which will supersede these study results.

6. STATISTICAL SOFTWARE AND QUALITY CONTROL

Day-to-day monitoring of the trial will be carried out by the Oxford Respiratory Trials Unit (ORTU) along with regular data queries and data checks by the trial statisticians.

Data is received by paper Case Report Forms (CRFs) from sites, entered into electronic database by Oxford. Data is held and managed by the ORTU.

All statistical analyses will be generated using SAS® software, version 9.4 or later. Graphics will either be generated using SAS or R software. Statistical analysis will be independently quality checked by a

second statistician to ensure that the SAP has been followed and programming errors are minimized.

7. PLANNED TABLES AND FIGURES

- Number of Subjects Randomized by Treatment Group and Site
- Patient Accountability Flow Chart, Overall and by Analysis Populations
- Reasons for Study Withdrawal by Treatment Group
- Protocol Deviations by Treatment Group
- Demographic and Baseline Characteristics by Treatment Group
- Comparability of Baseline Variables Across Sites
- Pool-ability analysis, Primary Endpoint Results by Site and Treatment Group
- Primary Endpoint Analyses
- Secondary Endpoint Analyses for CC with mild OSA per 2012 AASM Guidelines
- Secondary Endpoint Analyses for CC with mild OSA per 2007 AASM Guidelines
- Adverse Events by Treatment Group
- Secondary Endpoint Analyses for CC with moderate to severe OSA per 2007 AASM Guidelines
- Secondary Endpoint Analyses for CC with no OSA per 2007 AASM Guidelines but who enter the study per AASM 2012 Guidelines.
- Secondary Endpoint Analysis for Patients treated with CPAP comparing Patients Adherent and Non-Adherent to CPAP for CC with mild OSA per 2012 AASM Guidelines
- Secondary Endpoint Analysis for Patients treated with CPAP comparing Patients Adherent and Non-Adherent to CPAP for CC with mild OSA per 2007 AASM Guidelines
- Impact of Baseline Factors on Adherence to CPAP in Treatment Group
- AirView/CPAP Usage Data for Patients treated with CPAP, Overall and Comparing Patients Adherent and Non-Adherent to CPAP
- Number/Percent of Types of Contacts for Patients treated with CPAP, Overall and Comparing Patients Adherent and Non-Adherent to CPAP (+ types of interventions, problems)
- Primary Endpoint Analyses over Baseline AHI, Age, and BMI
- Secondary Endpoint (mean change in ESS and FSS) Analyses over Baseline AHI, Age, and BMI
- Association between Primary Endpoint and ODI and RERA Scores by Treatment Group and CPAP Adherence
- Baseline Demographic Characteristics, AHI, and Quality of Life Scores by Gender
- Secondary Endpoint Analyses by Treatment Group and Gender
- Summary of Change from Baseline in Weight and BMI by Treatment Group
- Primary Endpoint Summarized by Reasons for Patient Referral and Treatment Group

8. AMENDMENT

Amendment to SAP Version 1.4 **Effective as of: 19Jun2019**

Changes and additions to planned analyses

5.7 – 5.8 Primary and Secondary Endpoint Analyses
SAP version 1.4 specified a mixed-effects repeated measures model as the primary analysis for the

primary endpoint; the primary analysis population was the ITT-1 population. Sensitivity analyses of the primary endpoint, as well as all analyses of secondary endpoints, were also specified using the mixed-effects repeated measures model; the analysis populations of interest were the CC-1 and CC-2 (completed case) populations.

After reviewing preliminary results, the Trial Steering Committee determined that it would be more appropriate to examine endpoints in a true intent-to-treat fashion, including all patients in the analyses, rather than focusing on completed cases.

The primary analysis of the primary endpoint was performed as specified. For sensitivity and secondary endpoint analyses, the mixed-effects repeated measures analysis was replaced with analysis of covariance (ANCOVA) in the ITT-1 and ITT-2 analysis populations. The ANCOVA model adjusted for baseline score, where missing 3-month scores were replaced with baseline scores using a last observation carried forward (LOCF) approach. The advantages of using the ANCOVA LOCF model were (1) it was considered a conservative method of handling missing data, in which patients would have seen no improvement had they remained in the trial, and (2) this approach had the benefit of simplifying the underlying model assumptions.

Secondary endpoints were also analyzed using the ANCOVA LOCF model for the group of patients with no OSA (AHI<5) per **AASM 2007 guidelines** but who entered the study (AHI≥5) per **AASM 2012 guidelines**; these patients were considered the "very mild" group.

Additionally, effect sizes were calculated for each analysis model as

effect size = $\frac{treatment\ difference}{pooled\ SD}$

where

treatment difference = adjusted mean treatment difference (CPAP – SC) from analysis model

pooled SD = $SE_{trt\ diff}$ / $\sqrt{\left(\frac{1}{nCPAP}\right) + \left(\frac{1}{nSC}\right)}$ for mixed effects model

pooled SD = root MSE for ANCOVA LOCF model.

5.10 Additional Exploratory Analyses

Several of the exploratory analyses did not fit within the scope of the primary manuscript, and will be completed and reported at a later date:

- Summaries and analyses of secondary endpoints in patients with moderate OSA per AASM 2012 guidelines.
- Summaries and analyses of secondary endpoints comparing patients in the CPAP group who are adherent vs. not adherent to treatment.
- Summary of US Medicare definition of adherence.
- Predictors of adherence to CPAP therapy.
- Details of patient contacts.
- Summaries and analyses by gender.

9. DOCUMENT VERSION HISTORY

Version number Issue date	Author	Significant changes from previous version
1.0 28Sep2017	Gavin Reilly	NA
1.2 24Nov2017	Gavin Reilly	Updates to description of treatment groups, Number enrolled, compliance definition.
1.3 16Nov2018	Leslee Willes	SAP recreated following change in trial statisticians.
1.4 18Mar2019	Meredith Decker	Updates to subject recruitment dates, definition of HADS summary scores, analysis populations, exploratory analyses and definition of protocol deviations.
1.5 19Jun2019	Meredith Decker	Updates to the primary, secondary and exploratory analyses