ResMed Corp. Inogen Inc.

Portable Oxygen Concentrator Improvements to Physical Activity, Oxygen Usage, and Quality of Life in Chronic Obstructive Pulmonary Disease Patients using Long-term Oxygen Therapy (POC-STEP)

Study ID: MA-16-05-02
Protocol Version 1.0

Statistical Analysis Plan
Version: 1.0 – 18 Aug 2020

ResMed Corp.
9001 Spectrum Center Blvd.
San Diego CA 92123
(800) 424-0737

Inogen Inc.
326 Bollay Dr.
Goleta CA 93117
(800) 678-5572

CONFIDENTIAL **DO NOT COPY**

Table of Contents

1 Purpose of Statistical Analysis Plan

The purpose of this Statistical Analysis Plan (SAP) is to summarize the statistical methodologies being used to determine the required sample size for and to analyze all endpoints associated with the use of portable oxygen concentrators (POC) combined with standard of care (SOC) long-term oxygen therapy (LTOT) (treatment) versus SOC alone (control) for 12 weeks among subjects prescribed long term oxygen therapy for diagnosed COPD.

2 Early Study Closure

Due to the outbreak of 2019 Novel Coronavirus (COVID-19) in the United States in March 2020, in particular because of the inherent health risks in this for COPD patients on LTOT and the likely negative impact to their Physical Activity Levels (PAL) (i.e., the patient population and primary endpoint for this study, respectively), the decision was made to prematurely terminate the study effective April 30, 2020. At the time of study closure, 109 of the planned 190 subjects (57%) had been enrolled, of which only 3 were still actively participating in the study.

The following impacts of the COVID-19 outbreak from March 2020 through the final study site closure will be summarized in the final report, where applicable:

- A listing of all subjects impacted by COVID-19-related study disruption, including how they were impacted, along with any related analyses and/or discussions as appropriate
- Analyses and corresponding discussions that address the impacts of the implemented contingency measures (e.g., trial participant discontinuation from investigational product and/or study, alternative procedures used to collect critical safety and/or efficacy data) on the safety and efficacy results for the study
- A summary of the reasons for failing to collect any efficacy endpoints, including identifying specific limitation imposed by COVID-19 leading to the inability to perform the protocol-specified assessment

3 Study Design

3.1 Study Design/Overview

This is a post-approval, randomized, unblinded, multicenter, parallel group study with 12 weeks of follow-up. The study will be performed at up to twenty (20) centers in the United States. Up to 190 subjects (95 per treatment arm) will be enrolled in the study.

3.2 Device Descriptions

3.2.1 SOC Devices for Long-Term Oxygen Therapy

All subjects enrolled in the study will receive ambulatory and non-ambulatory SOC devices for long-term oxygen therapy such as oxygen tanks, transfill devices, and stationary oxygen concentrators.

3.2.2 POC (Investigational) Devices

Subjects randomized to the treatment arm will receive a commercially available POC to provide portable oxygen using pulsed-flow technology. Pulsed-dose oxygen is provided in a pre-programmed bolus amount only as the patient inhales. Each center will distribute either ResMed or Inogen POC devices, both of which have been cleared for market by FDA through the 510(k) process and are to be used in accordance with their respective current indications for use and User Manual/Operator's Manual. For the

purposes of this study, subjects randomized to the control arm may not use a POC device as an ambulatory oxygen option.

3.2.3 ActiGraph GT9X Link

All subjects enrolled in the study will receive an ActiGraph GT9X Link wearable device to collect physical activity data throughout the study. A CentrePoint (CP) data hub is also included with the ActiGraph device to upload the physical activity data.

3.2.4 Oxygen Use Diaries

All subjects enrolled in the study will receive a diary to record oxygen usage throughout the study. All subjects will record daily usage of SOC oxygen and subjects randomized to use a POC will also record daily usage of POC oxygen.

3.3 Study Schedule

3.3.1 Visit 1 (Enrollment/Baseline/Randomization)

Subjects will have an initial visit (Visit 1) to consent to the study, complete all baseline assessments, be randomized to a treatment group, and to receive their study devices (with training) and diaries.

3.3.2 Phone Calls (Week 1 and Week 6)

The centers will call subjects during Week 1 and Week 6 (± 3 days) of the study to asses and record updates on unreported or ongoing adverse events, record updates on respiratory condition-related concomitant medications, assess protocol compliance, address any questions or concerns, and provide additional device training (as needed).

3.3.3 Visit 2 (Week 12 Follow-Up)

Subjects will have a follow-up visit (Visit 2) during Week 12 (± 7 days) of the study to return all study devices and diaries, perform an ActiGraph device data download, complete all follow-up assessments, asses and record updates on unreported or ongoing adverse events, and record updates on respiratory condition-related concomitant medications. Completion of this visit will result in the completion of study participation

3.3.4 Randomization

Subjects will be randomized according to the pre-specified randomization schedule following the confirmation of enrollment eligibility. Subjects will be randomized to either the treatment or control arm in a 1:1 ratio. The randomization will be based on random, permuted blocks and stratified by center. Randomization will continue sequentially regardless of a subject not being treated according to their randomization assignment.

4 Study Objectives

This study is designed to evaluate the effect of using a POC combined with SOC long-term oxygen therapy versus SOC long-term oxygen therapy alone on the physical activity level of subjects with COPD requiring long-term oxygen therapy. This study will also assess the effect on oxygen usage, quality of life, hospitalizations, and death and the safety of the device.

5 Endpoints

5.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is defined as the change in physical activity level (PAL) from baseline to the week 12 follow-up visit.

5.2 Secondary Endpoints

The secondary endpoints are:

1. Average daily oxygen usage
2. Quality of life
   a. St. George's Respiratory Questionnaire
   b. Hospital Anxiety and Depression Scale
3. Six Minute Walk Test (6MWT) distance
4. Frequency of COPD exacerbation
5. COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
6. Average number of steps taken per day (non-hospital days)
7. Mortality due to COPD or other respiratory condition (primary cause of death)
8. All-cause mortality

5.3 Exploratory Endpoints

The exploratory endpoints are:

1. Time from enrollment to hospitalization for a COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
2. Number of days in hospital for COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
3. All-cause unplanned hospitalizations
4. Average daily activity duration and intensity
5. Average energy expenditure/day

5.4 Adverse Events

All adverse events will be recorded and reported.

6 Analysis of Primary Effectiveness Endpoint

The primary effectiveness endpoint analysis evaluates the difference in the mean change in PAL from baseline to the week 12 follow-up visit between the treatment and control arms. For accuracy, only wear-filtered ActiGraph data will be used in the analysis. The baseline PAL uses the average TEE and average SEE values during the first week of the study and the week 12 follow-up visit PAL uses the average values from the twelfth week of the study. Only subjects with available baseline and week 12 follow-up visit PAL values will be included in this analysis.

6.1 Hypotheses

The null hypothesis for the primary effectiveness endpoint is that there is no difference in the mean change in PAL for the treatment and control arms (indicating equality). Rejection of the null hypothesis

indicates the observed data supports the alternative hypothesis that a difference in the mean change in PAL exits between the treatment and control arms (indicating inequality). The hypotheses associated with the primary effectiveness endpoint are:

$$H_0: \mu_t - \mu_c = 0$$
$$H_1: \mu_t - \mu_c \neq 0$$

where μ_t and μ_c are the mean changes in PAL for the treatment and control arms, respectively.

6.2 Sample Size

The following assumptions were made while determining the required sample size:

- Patients will be assigned to either the treatment or control arm
- The primary analysis is an inequality comparison of the treatment and control arms
- The standard deviation for both arms is 0.22
- A difference ≥ 0.1 is considered clinically significant (i.e., minimal desired detectable difference)
- The difference between the arms is expected to be 0
- Statistical testing will use a two-sided 5% significance level with at least 80% power

At least 154 subjects (77 per treatment arm) are required to analyze the primary effectiveness endpoint with statistical power. To account for up to 20% attrition and other factors, up to 190 subjects (95 per treatment arm) are required to ensure adequate power is maintained for the primary effectiveness endpoint.

7 Analysis of Secondary Endpoints

The secondary endpoints are evaluated at the week 12 follow-up visit and compare the treatment and control arms. Analysis of the secondary endpoints evaluates the following:

1. Change in average daily oxygen usage from baseline
2. Change in Quality of Life from baseline
   a. St. George's Respiratory Questionnaire
   b. Hospital Anxiety and Depression Scale
3. Change in Six Minute Walk Test (6MWT) distance from baseline
4. Frequency of COPD exacerbation
5. COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
6. Change in average number of steps taken/day (non-hospital days) from baseline
7. Mortality due to COPD or other respiratory condition (primary cause of death)
8. All-cause mortality

For secondary endpoints 1 and 6, the baseline and week 12 follow-up visit values are the average value during the first and twelfth weeks of the study, respectively. For accuracy, only wake- and wear-filtered ActiGraph data will be used in the analysis.

The assessments used in secondary endpoints 2 and 3 are performed at both the baseline and week 12 follow-up visits.

7.1 Hypotheses

7.1.1 Continuous Secondary Endpoints

The null hypothesis for secondary endpoints 1-6 is that there is no difference in the endpoint means for the treatment and control arms (indicating equality). Rejection of the null hypothesis indicates the observed data supports the alternative hypothesis that a difference in the endpoint means exits between the treatment and control arms (indicating inequality). The hypotheses associated with these secondary endpoints are:

$$H_0: \mu_t - \mu_c = 0$$
$$H_1: \mu_t - \mu_c \neq 0$$

where μ_t and μ_c are the endpoint means for the treatment and control arms, respectively.

7.1.2 Categorical Secondary Endpoints

The null hypothesis for secondary endpoints 7 and 8 is that there is no difference in the endpoint proportions for the treatment and control arms (indicating equality). Rejection of the null hypothesis indicates the observed data supports the alternative hypothesis that a difference in the endpoint proportions exits between the treatment and control arms (indicating inequality). The hypotheses associated with these secondary endpoints are:

$$H_0: p_t - p_c = 0$$
$$H_1: p_t - p_c \neq 0$$

where p_t and p_c are the endpoint proportions for the treatment and control arms, respectively.

8 Analysis of Exploratory Endpoints

The exploratory endpoints are evaluated at the week 12 follow-up visit and compare the treatment and control arms. Analysis of the exploratory endpoints evaluates the following:

1. Time from enrollment to hospitalization for a COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
2. Number of days in hospital for COPD or other respiratory condition-related inpatient hospitalization ≥ 24 hours
3. All-cause unplanned hospitalizations
4. Change in average daily activity duration and intensity
5. Change in average energy expenditure/day

Exploratory endpoint 1 will be analyzed with a Cox proportional hazard model, using time to hospitalization as the response and treatment arm, center, age, and baseline 6MWT as predictors. The model will be summarized with regression coefficient estimates and standard deviations.

For exploratory endpoint 4, the baseline and week 12 follow-up visit values are the average value during the first and twelfth weeks of the study, respectively. For accuracy, only wake- and wear-filtered ActiGraph data will be used in the analysis. Exploratory endpoint 4 will be summarized in the following ways:

- Change in average duration per day at each intensity from baseline to the week 12 follow-up visit
- Average duration per day at each intensity over 12 weeks
- Change in percentage of time spent engaging in Sedentary to Very Low Physical Activity (SVLPA) from baseline to the week 12 follow-up visit

- Proportion of subjects engaging in Moderate to Vigorous Physical Activity (MVPA) at the week 12 follow-up visit who did not at baseline

For exploratory endpoint 5, the baseline and week 12 follow-up visit values are the average value during the first and twelfth weeks of the study, respectively. For accuracy, only wear-filtered ActiGraph data will be used in the analysis.

Exploratory endpoints 2-5 will only be summarized with descriptive statistics and will not have formal hypothesis testing performed.

9 Analysis of Adverse Events

All adverse events occurring during the study will be summarized by type, seriousness, Investigator-reported relatedness (to the study procedure and the LTOT, POC, and ActiGraph devices), severity, and outcome with the number of events experienced and the number and percentage of subjects experiencing the event. Adverse events will also be summarized in a listing. No formal hypothesis testing will be performed.

9.1 Seriousness

An adverse event is serious if it:

- Leads to death
- Leads to serious deterioration in the health of a patient that:
  - results in a life-threatening illness or injury,
  - results in permanent impairment of a body structure or body function,
  - requires inpatient hospitalization ≥ 24 hours or prolongation of existing hospitalization, or
  - results in medical or surgical intervention to prevent permanent impairment to a body structure or a body function.

9.2 Severity

Adverse events will be categorized by the Investigator as mild, moderate, or severe, depending on the event's impact on the subject's daily activity level:

- Mild: Usually transient, requiring no special treatment; does not interfere with the subject's daily activities.
- Moderate: Low-level inconvenience or concern to the subject; may interfere with daily activities, usually resolved by simple therapeutic non-interventional methods.
- Severe: Interruption in subject's daily activity requiring systemic drug therapy or other treatment.

9.3 Relatedness

Adverse events will be judged by the Investigator as to their relatedness to the study procedure and the LTOT, POC, and ActiGraph devices using the following classifications:

- Unrelated: The event is due to the underlying disease state or concomitant medication or therapy not related to the study-specific devices or procedures.
- Probably Not Related: The event had no significant temporal relationship to the study-specific devices or procedures and/or a more likely alternative etiology exists.

- Possibly Related: The event had a strong temporal relationship to the study-specific devices or procedures and alternative etiology is equally or less likely compared to the potential relationship to the study-specific devices or procedures.
- Probably Related: The event had a strong temporal relationship to the study-specific devices or procedures and another etiology is unlikely.
- Unknown: Relationship of the event to the study-specific devices or procedures and alternative etiology is unknown.

10 Measurements, Scales, and Subject-Reported Outcomes

10.1 ActiGraph Measurements

The ActiGraph GT9X Link wearable device continuously collects data including calories expended; duration of physical activity; X-, Y-, and Z-axis activity counts; steps taken; whether the subject is sleeping; and whether the device is being worn among many other measures. For the purposes of this study, only data recorded while the subject is wearing the device (wear-filtered) will be used. Certain measurements also only use data recorded while the subject is awake (wake-filtered) or asleep (sleep-filtered).

10.1.1 Total Energy Expenditure

The total energy expenditure (TEE) is the total number of wear-filtered kilocalories expended in a day.

10.1.2 Sleep Energy Expenditure

The sleep energy expenditure (SEE) is the total number of sleep- and wear-filtered kilocalories expended in a day.

10.1.3 Physical Activity Level

As it relates to the primary effectiveness endpoint, physical activity level (PAL) is the ratio of TEE to SEE, both as defined above, expressed as the following fraction:

$$\frac{\text{Total Energy Expenditure (TEE)}}{\text{Sleep Energy Expenditure (SEE)}}$$

10.1.4 Duration of Physical Activity

The duration of physical activity is the total number of wake- and wear-filtered minutes of physical activity performed in a day.

10.1.5 Intensity of Physical Activity

The intensity of the physical activity being performed is the categorization of the total X-, Y-, and Z-axis counts (vector magnitude unit or VMU) and the Y-axis count alone in a minute as one of the following:

- (SVLPA) Sedentary to Very Low: VMU < 3000
- (LMPA) Low to Moderate: VMU ≥ 3000 and Y-axis count < 1952
- (MVPA) Moderate to Vigorous: Y-axis count ≥ 1952

10.1.6 Steps Taken

The steps taken is the number of wear-filtered steps taken in a day.

10.2 Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) is a commonly used, validated questionnaire that measures anxiety and depression (Stern, 2014). The HADS consists of 14 questions about how the subject has been feeling over a preceding period, three months for the purposes of this study, with 7 pertaining to anxiety and 7 to depression.

Each question has four possible answers corresponding to a score between 0 and 3 that are added together to get a total score. The total scores for anxiety and depression are determined separately and are classified as one of the following:

- 0-7: Normal
- 8-10: Borderline abnormal (borderline case)
- 11-21: Abnormal (case)

10.3 Six Minute Walk Test

The Six-Minute Walk Test (6MWT), used in both practice and clinical research, assesses aerobic capacity and endurance. The 6MWT measures the distance (in meters) that can be walked on a flat, hard surface in six minutes.

10.4 St. George's Respiratory Questionnaire

The St. George's Respiratory Questionnaire (SGRQ) is a validated tool that can be used to measure health impairment in COPD patients. (Morishita-Katsu, et al., 2016). The SGRQ is a two-part questionnaire; Part 1 (questions 1-8) covers the patient's recollection of their symptoms over a preceding period, three months for the purposes of this study, and Part 2 (questions 9-16) addresses the patient's current state. The two parts are further categorized into three components: Symptoms (questions 1-8), Activity (questions 11 and 15), and Impacts (questions 9-10, 12-14, and 16-17).

Each possible answer for a question corresponds to a weight, between 0 and 100. The final score overall and for each component is the percentage of the total weight for selected answers out of the largest possible total weight for answers to the applicable questions. Higher final scores indicate more limitations.

11 General Statistical Considerations

11.1 Analysis Sets

11.1.1 Intention-to-Treat

The intention-to-treat analysis set (ITT) consists of randomized subjects who begin long-term oxygen therapy, with or without a POC, receive an ActiGraph device, and have at least baseline ActiGraph data. The ITT analysis set will be used as the primary analysis set in analyzing all endpoints.

11.1.2 Per Protocol

The per protocol analysis set (PP) consists of subjects in the ITT analysis set that did not experience any of the following major protocol deviations:

- Violations of inclusion/exclusion criteria for study
- Non-compliance (less than 18 hours of ActiGraph wear-time in a day) of more than two days during the first or twelfth week after enrollment

The PP analysis set will only be used, if different than the ITT analysis set, in a secondary analysis of the primary effectiveness endpoint assessing the sensitivity to the choice of analysis set.

11.2 Control of Systematic Bias

Several measures have been incorporated into the study design to minimize study including, but not limited to, the following:

- A multi-center trial minimizes investigator, center, or subject bias
- Consecutively eligible subjects should be enrolled into the study
- This Statistical Analysis Plan specifies appropriate statistical methodology to minimize bias
- Subjects are randomized to one of two arms, POC plus SOC for LTOT (treatment) or SOC for LTOT alone (control), which minimizes systematic bias in the subject treatment
- Standardized and validated questionnaires are used to collect data

11.3 Pooling Across Centers

All data will be pooled across centers for all analyses. Centers may be compared using summary statistics. An assessment of the appropriateness of pooling may be conducted using a linear regression model with the primary effectiveness endpoint as the outcome and center, treatment arm, and the interaction between the center and treatment arm as predictors.

11.4 Descriptive Statistics

Descriptive statistics will be presented for subject demographics, study disposition, screening information, ActiGraph device compliance, assessment results, and for any other characteristics data collected during the study.

11.5 Handling of Missing Data

A sensitivity analysis will be performed on the primary effectiveness endpoint to assess the effect of missing data. The results using the following imputation methods will be evaluated and compared to the results from the primary analysis:

- Missing week 12 follow-up visit PAL values are imputed with the most recent set of 7 consecutive days with available data from the second half of the 12-week follow-up period (i.e., weeks 7-12)
- Missing week 12 follow-up visit PAL values are imputed with the most recent set of 7 consecutive days with available data from the last quarter of the 12-week follow-up period (i.e., weeks 10-12)

11.6 Data Summaries

Unless otherwise stated, data will be summarized as followed:

- Continuous variables will be summarized with the count, mean, standard deviation, median, and range. Hypothesis testing will use the t-test (or Wilcoxon rank-sum test).
- Categorical variables will be summarized with counts and percentages. Hypothesis testing will use the chi-squared test (or Fisher exact test).
- Level of significance (α) for statistical testing will be 0.05 and two-sided.

11.7 Statistical Software

The statistical software package SAS® 9.4 or later will be used for all the data derivations, summarization, data listings and statistical analyses. Additional statistical software may be used for graphics or validation purposes as appropriate.

12 Version History

Version	Date	Description of Changes
1.0	18 Aug 2020	Initial description of analyses