



BTG

Interventional
Pulmonology

PneumRx

Elevair™ Endobronchial Coil System
Sponsor Executive Summary
June 14, 2018

Attachment 6: Proposed Post Approval Study Protocol

The table of contents can be found on page 2 of 24 or [click here](#).

Protocol CLN0009.p. F

**Lung Volume Reduction Coil for Treatment in Patients with Emphysema
(RENEW) Study**

**Statistical Analysis Plan
(Methodology)**

Version 3.0

November 30, 2015

Confidential and Proprietary

Nothing herein is to be disclosed without the written consent of PneumRx-BTG, Inc.

Table of Contents

1. SYNOPSIS OF STUDY DESIGN AND PROCEDURES	4
1.1 Study Objective	4
1.2 Study Design	4
1.2.1 Primary Effectiveness Endpoints.....	4
1.2.2 Secondary Effectiveness Endpoints.....	4
1.2.3 Other Effectiveness Endpoints.....	5
1.2.4 Safety Endpoints	5
1.3 Analysis Populations	6
1.3.1 Intent-to-Treat Population.....	6
1.3.2 Per-Protocol Population	6
1.3.3 Safety Population	6
1.4 Sample Size and Power Calculation	7
1.4.1 Primary Effectiveness Endpoint	7
1.4.2 Powered Secondary Effectiveness Endpoint.....	8
1.4.3 Power Calculation for Safety Endpoint	8
1.5 Randomization and Blinding.....	9
1.5.1 Randomization	9
1.5.2 Study Blind	9
1.6 Study Success.....	9
2. ANALYSIS CONSIDERATIONS	9
2.1 Subject Disposition	9
2.2 Demographic and Baseline Characteristics.....	9
2.3 Efficacy Endpoint Analyses.....	10
2.3.1 Primary Effectiveness Endpoint Analyses	10
2.3.2 Secondary Effectiveness Endpoint Analyses.....	11
2.3.3 Other Effectiveness Endpoint Analyses	12
2.3.4 Safety Endpoint Analyses	13
2.3.4.1 Event Time Periods.....	13
2.3.4.2 Safety Endpoint at 12 Months.....	13
2.3.4.3 Adverse Events	13
2.3.4.4 Serious Adverse Events	14
2.3.4.5 Re-hospitalization.....	14
2.4 Subgroup Analysis.....	15
2.6 Missing Data Handling	15
2.7 Poolability Issues	16
2.7.1 Multiple Center Effect.....	16

2.8 Interim Analysis 17

2.9 Sensitivity Analysis 17

2.10 Protocol Deviations 18

2.11 Additional Descriptive Analysis 18

2.12 Documentation and Other Considerations 19

3. REFERENCE 20

4. APPENDIX 22

4.1 Programming Specifications..... 22

1. SYNOPSIS OF STUDY DESIGN AND PROCEDURES

The purpose of the statistical analysis plan (SAP) is to provide a detailed and comprehensive description of the planned methodology and analysis for Protocol CLN0009.p. F, the RENEW Study. This plan is based on the version CLN0009.p. F, June 23rd, 2015 study protocol.

1.1 Study Objective

The primary objective of this study is to determine whether treatment with the RePneu LVRC[®] System results in improved exercise capacity and quality of life, as measured by improvements in the 6 Minute Walk Test (6MWT).

1.2 Study Design

This will be a prospective, multicenter, randomized, assessor-blinded controlled study comparing outcomes between the Lung Volume Reduction Coil (LVRC) and Control Groups. Subjects will be block randomized in the treatment (Lung Volume Reduction Coil, LVRC) to control group at a ratio of 1:1.

The randomization will be stratified by homogeneous versus heterogeneous emphysema, to support a balance of patients with differing heterogeneity in both the LVRC and Control groups per FDA's request.

There will be up to 315 subjects enrolled at up to 30 sites with 1:1 ratio of LVRC vs. Control, not including "roll-in" subjects.

1.2.1 Primary Effectiveness Endpoints

The primary effectiveness endpoint of the PneumRx RENEW study is the mean absolute change from baseline at 12 months in the 6 Minute Walk Test (6MWT), comparing LVRC and Control groups (overall type I error one-sided, $\alpha = 0.025$).

1.2.2 Secondary Effectiveness Endpoints

- Six Minute Walk Test (6MWT): responder analysis, comparing baseline to 12 months, LVRC vs. Control, responders defined as those with an improvement of ≥ 25 meters^[8]
- Forced Expiratory Volume in one second (FEV₁): mean percent change in FEV₁ results measured using spirometry, comparing baseline to 12 months, LVRC vs. Control

- St. George's Respiratory Questionnaire (SGRQ): mean absolute difference in SGRQ results comparing baseline to 12 months, LVRC vs. Control

1.2.3 Other Effectiveness Endpoints

- St. George's Respiratory Questionnaire (SGRQ): responder analysis, comparing baseline to 12 months, LVRC vs. Control, responders defined as those with an improvement of ≥ 4 points^[1]
- Residual Volume (RV): mean absolute difference in RV results measured using plethysmography, comparing baseline to 12 months, LVRC vs. Control^[2-4]
- Residual Volume/Total Lung Capacity (RV/TLC): mean absolute difference in RV/TLC results measured using plethysmography, comparing baseline to 12 months, LVRC vs. Control^[5-7]

1.2.4 Safety Endpoints

The safety analysis will tabulate the difference between the LVRC and Control groups in the proportion of subjects who experience one or more major complication(s) within 12 months post-baseline (and within defined blocks of time post-baseline). Major complications will be determined/adjudicated by the Clinical Events Committee.

Major Complications:

- Death;
- Pneumothorax that requires a chest drainage tube for more than 7 days (from time of chest drainage tube insertion to the time of chest drainage tube removal);
- Hemoptysis requiring blood transfusion(s), arterial embolization, or surgical/endoscopic procedure;
- COPD exacerbation that becomes life-threatening or disabling as a result of an increase in respiratory symptoms requiring in-patient hospitalization of >7 days with or without mechanical ventilation;
- Lower Respiratory Infections (including pneumonia) defined by new or increased clinical symptoms such as fever, chills, productive cough, chest pain, dyspnea and an infiltrate on plain chest x-ray and hospitalization for administration of intravenous antibiotics and/or steroids;
- Respiratory failure defined as a requirement for mechanical ventilatory support (whether via endotracheal tube or mask) for >24 hours; and

- An unanticipated bronchoscopy in order to remove one or more Coils due to a device-related AE. (Note: This definition does not include re-positioning, replacement or removal of the Coil(s) during the initial placement procedure.)

1.3 Analysis Populations

1.3.1 Intent-to-Treat Population

The Intent-to-Treat (ITT) population will include all subjects randomized to the LVRC or Control group, regardless of whether or not treatment was attempted.

Consenting subjects who withdraw consent prior to randomization, or who are found not to meet the inclusion/exclusion criteria prior to randomization, will be recorded on a screening log at each clinical site and will not be included in the ITT population.

1.3.2 Per-Protocol Population

The Per-Protocol (PP) population will include the subjects from the ITT population who complete the study without major study protocol deviations (PD) (i.e., any subject or Investigator activity that could have possibly interfered with the therapeutic administration of the treatment or the precise evaluation of treatment efficacy). The major study PDs are listed as follows:

- Informed Consent not obtained
- Inclusion/exclusion criteria not met
- 6MWT assessor blinding not maintained
- Pulmonary Function Test (PFT) assessor blinding not maintained
- Any of the major visits (such as visit 1, 2, 5, and 10) not done, with the exception of those who discontinued from the study due to (a) an adverse event related to the study device or procedure, or (b) a documented lack of treatment effect or (c) other reason;
- An alternative treatment, such as lung transplants, administered during the study

Other additional criteria may be added to the prior list to accommodate unforeseen events that occur during the conduct of the trial that result in noteworthy study protocol deviations. Any subjects excluded from the PP population will be identified and documented during the review of patient eligibility. Subjects in the PP group will be analyzed (grouped) by the actual treatment received.

1.3.3 Safety Population

The safety population will include all ITT subjects in the control group and all LVRC subjects who have at least one procedure done or who enter the procedure room, regardless of whether or not device deployment was attempted.

1.4 Sample Size and Power Calculation

There will be up to 315 subjects planned for this study, not including "roll-in" subjects. The sample size and power calculation is based on the primary and secondary effectiveness endpoint with overall type I error $\alpha = 0.025$ for a one-sided two-sample t-test.

For power estimates and sample size calculations for the current study design, PneumRx used historical OUS study data and the reported change in 6MWT for the control group of the Emphasys VENT Study.

1.4.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is the mean absolute change from baseline at 12 months in the 6 Minute Walk Test (6MWT), comparing LVRC and Control groups (overall type I error one-sided, $\alpha = 0.025$). It will be evaluated on both ITT and PP populations. The primary analysis will be based on the ITT population.

The hypothesis test is designed to show the significant improvement of the mean absolute change of 6MWT from baseline at 12 months in LVRC group comparing to Control group. The null (H_0) and alternative (H_1) hypotheses are:

$$\begin{aligned}H_0: \mu_T - \mu_C &\leq 0 \\H_1: \mu_T - \mu_C &> 0\end{aligned}$$

Where μ_T and μ_C equal the expected 12 month differences in 6MWT values from baseline for LVRC and control, respectively. The sample size calculation for the primary endpoint is based on the following assumptions:

- The true absolute mean change of 6MWT at 12 months in LVRC treatment group is assumed to be 49 meters (m) from PneumRx OUS clinical study protocol CLN0011. An estimated decline of 10 meters (-10m) of 6MWT is assumed to be the true absolute mean change of 6MWT at 12 months in Control group based on the VENT Study and PneumRx OUS clinical study protocol CLN0008. Therefore, the estimated difference between LVRC and Control groups is 59m.
- One-sided $\alpha = 0.025$
- A conservative estimate of 80m for the standard deviation is assumed based on PneumRx OUS study CLN0006

- 5% lost to follow up rate at 12 months

The power computations are based on the above estimates and one-sided superiority test at an alpha of 0.025 using nQuery Version 7.0 for a two-group t-test of equal means with unequal variance. A conservative estimate of 80m for the standard deviation showed that a sample size of 100 subjects per treatment arm would have power greater than 95%.

1.4.2 Powered Secondary Effectiveness Endpoint

Additional power calculation was conducted for the FEV1 secondary variable using the VENT trial. Specifically, let μ_T and μ_C represent the expected mean change of FEV₁ at 12 months from baseline for LVRC and Control, respectively. The null (H_0) and alternative (H_1) hypotheses for one-sided test are:

$$H_0: \mu_T - \mu_C \leq 0$$

$$H_1: \mu_T - \mu_C > 0$$

The estimates of change from baseline at the endpoint visit were 0.06 and 0.01 for the VENT and control groups, respectively. With a standard deviation of 0.10 and one-sided t-test of $\alpha = 0.025$ and assuming 5% lost to follow-up at 12 months, a sample size of 151 per treatment arm will have a 95% power to detect a difference between LVRC and Control groups. It is expected that the power for percent change will be similar to absolute change.

The sample size and power calculations were performed using nQuery Version 7.0.

1.4.3 Power Calculation for Safety Endpoint

For the proportion of subjects experiencing major complications, 151 subjects per treatment group would provide approximately 80% power to detect a 12% difference between treatment groups.

To summarize, approximately 151 subjects per treatment arm will be needed in the study due to the considerations of power for the primary, secondary and safety endpoint tests. With additional consideration of 5% subjects who may not be evaluable, approximately 158 subjects per treatment arm (315 subjects total) are planned for enrollment in this study.

Note that roll-in subjects are planned for this study in addition to the study estimate of 315 subjects to be enrolled for hypothesis testing. Roll-in subject data will not be included in these analyses, but will be reported separately.

1.5 Randomization and Blinding

1.5.1 Randomization

Subjects will be block randomized in LVRC (Treatment) to Control groups at the ratio of 1:1. The randomization will be stratified by homogeneous versus heterogeneous emphysema, to support a balance of patients with differing heterogeneity in both the LVRC and Control Groups per FDA's request.

1.5.2 Study Blind

The subject as well as the investigator performing the procedure will not be blind to the study treatment. The investigator will however not assess the subjects for the effectiveness endpoints (pulmonary function tests (PFT) and 6MWT). The PFT and 6MWT assessor (i.e. those working with the subject to collect data on the 6MWT, SGRQ, plethysmography measures, and spirometry measures) will be blinded to the treatment received by the subject and the subjects will be instructed not to share any information that may identify the treatment received with the assessors.

1.6 Study Success

For the trial to be successful, the mean improvement in 6MWT from the first treatment visit to 12 months must demonstrate a statistically significant difference between LVRC and Control groups. Additional labeling claims may be made based on the secondary endpoints.

2. ANALYSIS CONSIDERATIONS

2.1 Subject Disposition

The number of subjects screened, randomized and treated in the study will be presented by treatment group overall and by analysis population. Also presented will be the number and percentage of subjects who attend each of the study visits, who complete the study, and who terminate the study prematurely, overall and by their reasons for premature termination. Subjects excluded from any analyses will be summarized with reasons for exclusion.

2.2 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized by treatment group for the ITT population, safety population, and the PP population. For all assessments the baseline value will be defined as the last assessment evaluated prior to treatment, e.g., 'Pre-treatment screening', or visit

1. Demographic and baseline characteristics will be presented with summary statistics (sample size (N), mean, standard deviation (STD), median, minimum, and maximum) for continuous variables and frequency distributions for categorical variables. These characteristics will also be summarized by treatment group for the ITT population, safety population, and the PP population. For continuous variables (e.g., age), comparisons between the two treatment groups will be conducted using a two-way analysis of variance with factors of treatment group and investigational site. Frequencies and percentages of races will be presented by treatment group. Other discrete variables (e.g., gender) will be summarized using frequencies and percentages; the treatment groups will be compared using the Cochran-Mantel-Haenszel (CMH) test stratified by investigational site. Past and current medical conditions, medical history and ethnic origin, will not be compared statistically.

2.3 Efficacy Endpoint Analyses

The hypothesis testing for primary effectiveness endpoint will be one-sided at $\alpha = 0.025$ significance level. The hypothesis testing for the secondary endpoints will be one-sided with adjustments on family wise type I error at $\alpha = 0.025$, using the Hochberg step-up procedure. Efficacy analyses will be performed for both the ITT and PP populations, with the primary analysis based on the ITT population.

2.3.1 Primary Effectiveness Endpoint Analyses

The primary effectiveness endpoint, change in 6MWT from Baseline (Pre-Treatment Visit) to the 12 month Follow-Up Visit, will be expressed as a mean absolute change in meters. The statistical testing will be based on an analysis of covariance (ANCOVA) with factors of treatment and analysis center and covariates of baseline 6MWT and emphysema heterogeneity. The inclusion of the emphysema heterogeneity as a covariate (homogeneous emphysema or heterogeneous emphysema) was at the FDA's request, albeit no statistically different outcomes were noted in OUS data, and such data were previously reported to FDA. Additionally, the statistical significance of the treatment by analysis center will be evaluated to assess the appropriateness of pooling the data across centers (see section 2.7).

The statistical hypotheses notation follows. Let μ_T and μ_C equal the expected 12 month difference in 6MWT values from baseline for LVRC and Control, respectively. The null and alternative one-sided hypotheses are:

$$H_0: \mu_T - \mu_C \leq 0$$

$$H_1: \mu_T - \mu_C > 0$$

The testing of these hypotheses will use the estimates derived from the ANCOVA described above with H_0 being rejected if the confidence interval for $\mu_T - \mu_C$ is greater than 0. The confidence coefficient will be 97.5%.

Tests of superiority will be based on either parametric or non-parametric methods consistent with the statistical assumptions required to support the analyses. Specifically, the tests of superiority will be based on an ANCOVA with factors of treatment and analysis center (see section 2.7) and baseline 6MWT and emphysema heterogeneity as covariates or on ranked data submitted to an ANCOVA with factors of treatment and analysis center and baseline 6MWT and emphysema heterogeneity as covariates. An evaluation of the residuals from ANCOVA based on 6MWT for complete cases will be performed both graphically and quantitatively to assess the normality assumption. If distributions are markedly skewed, a rank transformed ANCOVA analysis will be conducted. Results of the rank-transformed analyses then will be considered the primary analysis; however, results of the non-ranked-transformed analyses will also be presented.

2.3.2 Secondary Effectiveness Endpoint Analyses

All secondary effectiveness analyses will be performed on both ITT and PP populations, with one-sided tests at $\alpha = 0.025$. Appropriate adjustments will be made to the tests of these secondary endpoints to account for the impact on Type I error using the step-up methods described by Hochberg (1988)^[13].

The following secondary endpoints will be tested for their statistical significance:

- 6MWT: responder analysis, comparing baseline to 12 months, LVRC vs. Control, responders defined as those with an improvement of ≥ 25 meters^[8]
- FEV₁: mean percent change (12 month value minus baseline value) in FEV₁ results measured using spirometry, LVRC vs. Control
- SGRQ: mean absolute difference (baseline score minus 12 month score) in SGRQ results, LVRC vs. Control

The proportion of 6MWT responders will be compared using logistic regression with factors of treatment and analysis center and covariates of baseline 6MWT and emphysema heterogeneity. A subject will be classified as a 6MWT responder if their 12 month change from baseline in 6MWT is at least 25 meters.

Let π_T and π_C equal the expected proportion of 6MWT responders, and θ_T equal the expected odds of 6MWT responders ($\pi_T/(1 - \pi_T)$). The null and alternative one-sided hypotheses are:

$$H_0: \log \theta_T - \log \theta_C \leq 0$$

$$H_A: \log \theta_T - \log \theta_C > 0$$

The testing of these hypotheses will use logistic regression analysis described above with H_0 being rejected if the one-sided confidence interval for $\log \theta_T - \log \theta_C$ is greater than 0, or equivalently, if the confidence interval for the odds ratio, θ_T/θ_C , is greater than 1. The confidence coefficient will be 97.5%. Estimates and confidence intervals for the proportions π_T and π_C will also be presented.

For FEV₁ and SGRQ continuous secondary endpoints, the inferential p-values comparing the two groups will be computed following the same methodology specified for the primary variable, using the appropriate baseline values as covariates.

Specifically let μ_T and μ_C equal the expected 12 month difference in SGRQ or percent difference in FEV₁ from baseline for LVRC and Control, respectively. The null and alternative one-sided hypotheses are stated as:

$$H_0: \mu_T - \mu_C \leq 0$$

$$H_1: \mu_T - \mu_C > 0$$

The testing of these hypotheses will use the estimates derived from the ANCOVA (or ANCOVA of ranks in the presence of significant skewness) described above for the primary endpoint, with H_0 being rejected if the confidence interval for $\mu_T - \mu_C$ is greater than 0. The confidence coefficient will be 97.5%.

2.3.3 Other Effectiveness Endpoint Analyses

Other effectiveness endpoints will also be tested for their statistical significance on both ITT and PP populations, with one-sided tests at $\alpha = 0.025$, without adjustment for multiplicity, using the same methodology as for the key secondary endpoints and using the corresponding covariate variable(s).

- St. George's Respiratory Questionnaire (SGRQ): responder analysis comparing baseline to 12 months, LVRC vs. Control, responders defined as those with an improvement of ≥ 4 points^[1]
- Residual Volume (RV): absolute difference in RV results measured using plethysmography, comparing baseline to 12 months, LVRC vs. Control^[2-4]

- Residual Volume/Total Lung Capacity (RV/TLC): absolute difference in RV/TLC results measured using plethysmography, comparing baseline to 12 months, LVRC vs. Control^[5-7]

2.3.4 Safety Endpoint Analyses

Treatment emergent AEs and major complications will be summarized by treatment group for the safety population. AE and major complication summaries will include all events experienced within the first 12 month visit. Overall summaries will be presented, but shorter periods of time prior to 12 month visit will be considered as well. In order to assess the safety profile of the LVRC without the interference of the procedure effect, the timing of adverse events and major complications with respect to the treatment visits will be considered.

2.3.4.1 Event Time Periods

In order to assess the safety profile of the LVRC without the interference of the procedure effect, adverse events and major complications will be summarized by time period. Each event will be included in the time period in which the event began, with the periods defined as follows:

- Peri-procedural: 0-30 days post each of the two treatment visits (visit 2 or visit 5)
- Between treatment #1 and visit 5: More than 30 days after visit 2, but prior to visit 5
- Between treatment #2 (or treatment #1 for subjects that miss Visit 5) and 9 months: More than 30 days after visit 5, but prior to 9 months post treatment #1
- Between 9 and 12 months: 9 or more months post treatment #1 through visit 10

Regarding the programming specifications for inclusion of AE records through the 12-month visit, please refer to Appendix 4.1 (item F).

2.3.4.2 Safety Endpoint at 12 Months

Major complications at 12 months will be summarized by treatment group with both event and subject counts. Additionally, major complication event rates will be presented. Event rates will be computed using Poisson regression so each subject's follow-up time will be considered along with event counts.

The proportion of subjects in each treatment group who experience one or more major complication(s) will be reported along with exact 95% confidence intervals. A statistical comparison between the proportions of subjects in each treatment group will be evaluated with Fisher's exact test.

2.3.4.3 Adverse Events

AE characteristics, including severity, seriousness, and relationship to the device and/or procedure will be presented in listings, and AE counts will be presented per both event and subject levels by treatment group.

MedDRA system organ classes and preferred terms will be summarized for AE categories by treatment group. Summaries will include event and subject counts, and Poisson regression event rates. For the events summarized by subject counts, each subject will be counted only once within a system organ class or a preferred term by using the adverse events with the highest severity within each category. For system organ classes and preferred terms with an incidence of more than five percent in either treatment group, statistical comparisons between treatment groups will be evaluated with Fisher's exact test. Event rates will be computed using Poisson regression so each subject's follow-up time will be considered along with event counts.

In addition, all information pertaining to AEs noted during the study will also be listed by subject. Details of the line listing by subject will include verbatim term given by the Investigator, preferred term, system organ class, start date, stop date, severity, and device or procedure relatedness. The AE onset will also be shown relative (in number of days) to the day of the procedure.

2.3.4.4 Serious Adverse Events

The incidence and severity of serious adverse events (SAEs) will be tabulated by subject. In addition, a list of subjects who were discontinued from the study will also be provided. The standard definition of a Serious Adverse Event (SAE) per 21 CFR §§812 and 803 will be followed.

Overall summary tables of SAE characteristics and event and subject counts of MedDRA system organ classes and preferred terms will be presented for all SAEs. Summaries and analyses will be similar to those performed for AEs including between treatment comparisons when incidence is greater than five percent in either treatment group.

2.3.4.5 Re-hospitalization

Re-hospitalization rates will be reported by treatment group on a Per-Subject basis and on a Per-Event basis. The Per-Subject re-hospitalization rate is the proportion of subjects who were re-admitted post-discharge. An individual subject will only be counted once in the Per-Subject no matter how many times they are readmitted during the follow-up period. The Per-Event re-hospitalization rate is the proportion of hospital re-admissions per treatment group including multiple re-admissions per individual subject as determined using a Poisson regression model, including an offset parameter for each patient to represent the follow-up duration post-discharge during which hospitalizations were recorded. These data will be summarized by treatment group.

2.4 Subgroup Analysis

The following subgroup analyses will be presented for the primary and secondary endpoints using ANCOVA or logistic regression after MCMC multiple imputation with a factor of treatment and the corresponding baseline value and emphysema heterogeneity as covariates for the ITT population:

- US vs. OUS (outside of the US)
- Heterogeneity of emphysema
- Severity of air trapping ($RV \geq 225\%$ vs. $RV < 225\%$)
- Gender

2.6 Missing Data Handling

The extent and pattern of missing data for primary and secondary efficacy endpoints will be summarized separately by treatment group. All missing 12 month values for primary and secondary efficacy endpoints will be estimated by multiple imputation. The following procedure describes the analysis process based on Markov Chain Monte Carlo (MCMC) multiple imputation for continuous variables.

The following procedure will be used for the 6MWT, percent change in FEV1 and the SGRQ. The least squares mean and standard error for the change from baseline in 12 month analysis values will be derived from an analysis using the method of Markov Chain Monte Carlo (MCMC) multiple imputation^[14-18]. Multiple imputation will involve 4 distinct phases or, using Rubin's (1987)^[19] terminology, tasks:

1. The number of missing values to be estimated by MCMC (nmiss) for 12 month value will be calculated.
2. Create a data set of subjects with observed actual values and those needing estimation by MCMC. The missing values in the data set will be filled in using the MCMC method 50 times to generate 50 data sets.
3. For each complete data set, the least squares mean difference in change from baseline to 12 months for LVRC minus the Control and its standard error will be calculated using an analysis of covariance (ANCOVA) with factors of treatment group and analysis center and covariates of baseline parameter analyzed and emphysema heterogeneity.

4. The results from these analyses will be combined into a mean between treatment difference and standard error in change from baseline to 12 months using SAS PROC MIANALYZE. The final confidence interval for the LVRC minus Control group will be derived from this mean difference and standard error.

A similar procedure will be used for the analyses based on proportion of 6MWT responders at 12 months wherein the ANCOVA analysis is replaced with a logistic regression. Specifically, missing dichotomous values for 6MWT responder status will be calculated from the 12 month values estimated by (MCMC). The overall log odds of responders and percent of responders at 12 months with its corresponding standard error will be derived from an analysis using the method of Markov Chain Monte Carlo (MCMC) multiple imputation^[14-18]. Multiple imputation will involve 4 distinct phases or, using Rubin's (1987)^[19] terminology, tasks:

1. For each complete data set with 12 month missing values filled in using the MCMC method above, the percent of responders and the difference in log odds of responders at 12 months between treatments will be calculated using a logistic regression with a factor of treatment group and analysis center and covariates of baseline parameter analyzed and emphysema heterogeneity.
2. The results from these analyses will be combined into an overall percent of responders and difference in log odds of responders at 12 months between treatments with a corresponding standard error for each treatment group using SAS PROC MIANALYZE. The final confidence interval for the difference in log odds of responders for the LVRC and control will be derived from these estimates.

2.7 Poolability Issues

Because this is a multi-center study, analysis will be performed by pooling data across study sites. Therefore, the treatment effect by investigational site will be evaluated to assess the appropriateness of pooling the data across centers^[20-22].

2.7.1 Multiple Center Effect

The clinical study will be conducted under a common protocol for each investigational site with the intention of pooling the data for analysis. Every effort will be made to promote consistency in study execution at each study site. The study is expected to be conducted with a minimum of 5 subjects randomized and included in each treatment group in the ITT population for each investigational site. In the event that there are too few subjects in a treatment group for an investigational site, then

the site's data will be combined with other site(s) to achieve the desired sample size minimum per treatment group.

The combining of site data will be accomplished by ranking the sites that did not enroll at least 5 subjects per treatment group by total enrollment, with ties broken by site number. Among those sites, the site with the largest enrollment will be combined with as many of the smallest sites as necessary until the minimum enrollment requirement is met. The process is repeated among the remaining sites that did not meet the required enrollment. Investigational sites will be combined within geographic region only (US vs OUS).

The process of combining investigator data that have insufficient subjects per treatment group will result in re-defining the groups of investigators for the purposes of statistical analyses. These combined groups will be referred to as "analysis centers" in the statistical analyses.

Prior to multiple imputation, the consistency of treatment response will be investigated across the analysis centers subsequent to combining the data as described above. For the purpose of testing consistency of treatment response, the primary efficacy variable will be considered. 6MWT will be analyzed with an ANCOVA (ranked or unranked based on the assessment of skewness per Section 2.3.1) with factors of treatment group, analysis center and treatment by analysis center interaction, and baseline 6MWT and emphysema heterogeneity as covariates. Estimates of treatment effect and CIs will be calculated separately by analysis center. Further examination via graphical means and/or sensitivity analyses will follow if the primary endpoint analysis results in a significant interaction term, to assess whether the nature of the interaction is quantitative or qualitative and whether there are extreme analysis centers that could affect the interpretation of common statistical and clinical conclusions.

2.8 Interim Analysis

The efficacy and safety analyses as described herein will occur following the completion of the 12-month follow-up visit for all patients per the protocol defined endpoints. Although this is an ongoing study with 60-month follow-up, there are no formal additional interim analyses planned for this study. Analyses of the follow-up period subsequent to the 12-month visit will be detailed in a separate statistical analysis plan.

Interim study reports with descriptive safety analysis may be produced per regulatory or Data Monitoring Committee (DMC) charter and requests.

2.9 Sensitivity Analysis

The following sensitivity analyses will be conducted to explore the impact of missing observation estimation on efficacy assessment of 6MWT.

- Multiple Imputation with SAS PROC MI 6MWT values at post-procedure visits will be imputed using a parametric regression model with covariates of baseline 6MWT, and emphysema heterogeneity as well as the assumption of multivariate normality and a monotone missing data pattern.
- Generalized Estimating Equations (GEE) models will be used to analyze the repeated 6MWT values. The GEE method uses all available data from all subjects and accounts for both the within and between subject sources of variation in the repeated measures over time. A first-order auto-regressive correlation structure will be used. Parameters for time, treatment group, and analysis center and covariates of baseline 6MWT and emphysema heterogeneity will be included in the model.
- 'Complete case' analysis will be performed using ANCOVA with factors of treatment and analysis center and baseline 6MWT and emphysema heterogeneity as covariates.
- Worst case sensitivity analysis classifies all missing 12-month 6MWT data as a "Failure" (using baseline 6MWT or their last observation of 6MWT, whichever is worse, for the 12-month 6MWT).

2.10 Protocol Deviations

For each treatment group, the frequency of protocol deviations (PD) will be summarized by major and minor, by center, by category and by category within major and minor.

2.11 Additional Descriptive Analysis

Additional descriptive summary statistics will be presented for clinical assessments (such as plethysmography and spirometry measures), procedural and device data, and efficacy results.

For continuous variables, summary statistics (sample size (N), mean, standard deviation (STD), median, minimum, and maximum), and 95% CIs will be presented by treatment group and by visit, where applicable.

For categorical variables, results will be summarized by treatment group and by visit with subject counts and percentages/rates, as well as the exact 95% Clopper-Pearson^[10] CIs.

For both continuous and categorical variables, p-value will be provided for the comparison between two treatment groups.

Additionally, descriptive summaries of other responder rates for treatment efficacy at 12 months, such as 6-minute walk distance ≥ 54 meters, FEV₁ improvement $\geq 12\%$, RV change ≥ 0.35 , or SGRQ score change ≥ 8 points will be presented. Survival analysis will be performed to compare the time to death event between LVRC and Control groups with log-rank test. Survival curves will be constructed using Kaplan-Meier estimates in order to analyze the survival distributions through 30

days, 180 days, 270 days, and 365 days post treatment #1 for death event. Subjects without events will be censored at their last known event-free time point.

2.12 Documentation and Other Considerations

All analyses will be performed using SAS[®] for Windows, version 9.3 and above.

3. REFERENCE

1. Schünemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbings D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56: 1170-1176
2. O'Donnell, DE et al. Effects of tiotropium on lung hyperinflation, dyspnea and exercise tolerance in COPD. *Eur. Respir. J* 2004; 23; 832-40.
3. O'Donnell, DE et al. Effects of Fluticasone Propionate/Salmeterol on Lung Hyperinflation and Exercise Endurance in COPD. *CHEST*; 130; 3; Sept. 2006; 647-56 .
4. O'Donnell, DE et al. Hyperinflation, dyspnea, and exercise intolerance in chronic obstructive pulmonary disease. *Proc Am Thorac Soc.* 2006 Apr; 3(2):180-4. Review.
5. Tzani, P et al. Effects of beclomethasone/formoterol fixed combination on lung hyperinflation and dyspnea in COPD patients. *CHEST*; 130; 3; Sept. 2006; 647-56.
6. Puhan MA, Chandra D, Mosenifar Z, Ries A, Make B, Hansel NN, Wise R, Sciruba B for the National Emphysema Treatment Trial (NETT) Research Group. The minimal important difference of exercise tests in severe COPD. *Eur Respir J.* 2011; 27: 784-90.
7. Celli, B. et al. Improvements in Resting Inspiratory Capacity and Hyperinflation with Tiotropium in COPD Patients with Increased Static Lung Volumes. *CHEST*; 124; 5; Nov. 2003; 1743-48.
8. Holland AE, Hill CJ, Rasekaba T, Lee A, Naughton MT, McDonald CF. Updating the minimal important difference for six-minute walk distance in patients with chronic obstructive pulmonary disease. *Arch Phys Med Rehabil* 2010; 91: 221-225.
9. Marek Fisz, Probability Theory and Mathematical Statistics (3 edition, ISBN-13: 978-0898741797), June 1980.
10. Clopper CJ, Pearson E. The Use of the Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* 1934; 26:404-413.
11. Newcombe R. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 1998; 18: 873-890.
12. Cochran, W. G. The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 25, 315-345, (1952).
13. Y, Hochberg. 1987. *Multiple Comparison Procedures*. New York: John Wiley and Sons
14. Barnes SA, Lindborg SR, Seaman JW. Multiple Imputation Techniques in Small Sample Clinical Trials. *Stat Med* 2006 Jan 30; 25(2):233-45.
15. Lavori PW, Dawson R, Shera D. A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data. *Statistics in Medicine* 1995; 14:1913-1925.
16. Li X, Mehrotra DV, Barnard J. Analysis of Incomplete Longitudinal Binary Data Using Multiple Imputation. *Stat Med.* 2006 Jun 30; 25(12):2107-24.
17. PD., Allison. Multiple Imputation for Missing Data. *Sociological Methods & Research*, Feb 2000; Vol. 28 No. 3:301-309.
18. Little RJA, Schenker N. Handbook of Statistical Methodology-Missing Data. New York: Plenum Press; 1995.
19. Little RJA, Rubin DB. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons;
20. JL., Fleiss. Analysis of data from multiclinic trials. *Control Clin Trials* 1986; 7:267-75.

21. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001; 135:112-23.
22. S., Senn. Some controversies in planning and analyzing multi-centre trials. *Stat Med* 1998; 17:1753-65.
23. Zar, J. H. (1984). *Biostatistical Analysis*. London: Prentice-Hall International.

4. APPENDIX

4.1 Programming Specifications

- A. Required margins: at least 1.25 inches on the binding margin and at least 1 inch on all other sides. All output should have the following header at the upper left margin:

PneumRx - BTG International
Protocol: CLN0009.p. F

and the following header (right-justified) at the upper right margin:

Page n of N
DDMMMYYYY

Tables/appendices/listings should be internally paginated (i.e., page numbers should appear sequentially within each table).

All output should have SAS program name in the lower right and the data source(s) used to generate the output in the lower left:

Data

Source: xxx

PROGRAM: XYZ.sas

- B. In general, data listings should be sorted by treatment group, subject number and visit/start dates, unless specific instructions to do otherwise.
- C. The following algorithm should be used to impute adverse event start dates for which only partial information is known:
- Missing day and month
 - If the year is same as the year of treatment(LVRC) or Visit 2(Control), then the day and month of treatment (LVRC) or Visit 2 (Control) is assigned to the missing fields.
 - If the year is prior to or after the year of treatment, then January 1 is assigned to the missing fields.
 - Missing month only
 - Treat day as missing and replace both month and day according to the above procedure.
 - Missing day only
 - Then the first day of the month is assigned to the missing day.

If the AE date of resolution is complete and the imputed start date as above is after the resolution date, the start date is imputed using the resolution date.

Adverse events with partially missing stop dates are imputed a resolution date as follows:

- *year is missing* – the date is left missing.
- *month is missing* – impute “December.”
- *day is missing* – impute last date of that month.

D. Complete dates for concomitant medications (CM) with missing or partially missing start dates are imputed using the same algorithm described for adverse event onset dates. If the end date is missing or partially missing, the imputation rule is applied in the following order:

- 1) *year is missing* – the medication is considered to have been received at all periods after the period determined by the start date. The date is left missing.
- 2) *month is missing* – impute “December.”
- 3) *day is missing* – impute last date of that month.

E. Date imputations are applied to the process of assigning treatment period and study day and should be retained in the derived database, but the data listings should display the original, partially missing dates.

F. Specifications regarding the inclusion of AE/CM/PD records up to 12-month visit are defined as follows:

- LVRC Arm
 - a) Completed 12 month visit – then include all records with AE/CM/PD onset dates \leq the actual 12 month visit date.
 - b) Missing 12 month visit – then include all records with AE/CM/PD onset dates \leq 13 months post treatment #1 date.
- Control Arm
 - a) Completed 12 month visit – then include all records with AE/CM/PD onset dates \leq the actual 12 month visit date.
 - b) Missing 12 month visit – then include all records with AE/CM/PD onset dates \leq 13 months post Visit 2 date.

G. Unless otherwise noted, the mean (standard deviation) of a set of values should be printed out to one (two) more decimal(s) than the raw value.

e.g., raw: xx
mean and standard deviation: xx.x and xx.xx
range (minimum and maximum): xx, xx

- H. All table percentages should be reported with one decimal point unless otherwise noted.
- I. Missing data should be represented on patient listings as 1) dashes “-,” and properly footnoted: “- = data not available” or 2) “n/a,” with footnote “n/a = not applicable,” whichever is appropriate.
- J. Times should be printed in the format “HH:MM.” “HH” represents the 2-digit hour portion of the time. “MM” represents the 2-digit minute portion of the time. Both hour and minute portions of time are zero-filled on the left if they have only one digit. Missing time portions should be represented on patient listings as dashes (“10:--“ and “--:--”). Times that are missing because they are not applicable for the patient should be printed as “n/a,” unless otherwise specified.