Endobronchial Valve for Emphysema Palliation Trial (VENT):

Statistical Analysis Plan

Analysis Summary

Study Title: Endobronchial Valve for Emphysema Palliation Trial (VENT) (Clinical Protocol 630-0001 - J).

Primary Objectives: (1) To determine whether the mean FEV_1 AND mean distance traveled in the six minute walk test (6MWT) will be improved in patients treated with the Emphasys Endobronchial Valve (EBV) compared to randomized controls receiving optimal medical management at 180 days after treatment. (2) To demonstrate that the rate of major complications at 180 days after treatment in patients treated with the EBV is acceptable.

Secondary Objectives: (1) To identify patient characteristics that may affect the outcome of the EBV procedure and to shed further light on the mechanisms of improvement with the EBV procedure. (2) To compare the response between treatment arms of the following secondary endpoints: Mean change in QOL (as measured by St. George's Respiratory Questionnaire (SGRQ), Modified Medical Research Council Rating of Dyspnea Score (mMRC), maximal work load as measured by cycle ergometry, and an assessment of room oxygen requirement. Appropriate adjustments will be made to account for impact on Type I error for these secondary endpoints.

Additional analysis will be performed on the following: mean change in RV, DL_{CO}, Quality of Well Being (QWB) scale and Body Mass Index – Obstruction Metric – Dyspnea Score – Exercise Tolerance Composite (BODE) index. Further the following variables will be evaluated: per-patient success rate of clinically significant improvement/maintenance in FEV_1 and 6MWT, procedure success rate (EBV only), technical success rate (EBV only), complications, device-related Adverse Events during procedure, re-hospitalization, device-related Adverse Events post discharge, survival, and freedom from death, lung transplantation, and LVRS.

In light of the National Emphysema Treatment Trial (NETT) results showing that exercise tolerance and disease distribution are covariate predictors of outcomes with lung volume reduction surgery (LVRS), VENT will stratify in a similar fashion to evaluate this relationship in the context of EBV.

Overall Definition of Success

The study will be a success if:

For effectiveness the differences between arms for the mean percent change from baseline at 180 days for both FEV_1 AND 6MWT reach statistical significance (one-sided test at $p<0.025$) in favor of the treatment group, and for safety the 95% upper confidence interval for the Major Complication Composite (MCC) rate delta between the EBV arm and the control arm is $\leq$ 30%.

Patient Population: Male and female patients with heterogeneous emphysema as defined in the inclusion/exclusion criteria and radiological inclusion criteria.

Study Design: A multicenter, randomized trial of the Transcopic Emphasys Endobronchial Valve (TS-EBV) compared to optimal medical management controls. The randomization ratio will be two EBV patients to one control. Patients will be followed for one year.

Number of Patients: A minimum of 246 patients (a minimum of 164 treated with the TS-EBV).

Statistical Procedures

I. Introduction - Emphysema is characterized by the gradual, irreversible breakdown of tissue and loss of elastic recoil within the lungs, causing them to lose the ability to expel air and efficiently absorb oxygen. As this chronic condition inexorably progresses, the diseased, hyperinflated areas of the lung eventually fill the chest cavity, leaving less and less volume available for the viable lung tissue. Emphasys has developed a bronchoscopic approach to block the inspiratory airflow into targeted, hyperinflated regions of the lung, while permitting exhaled gas to escape. Emphasys Medical theorized that if a one-way valve was utilized to block inspiratory airflow, there would be a vent for collateral airflow and mucus. This approach may provide some of the clinical benefits of LVRS without the high risks and costs associated with such an invasive surgical procedure.

II. Characteristics of Study Variables and Descriptive Data –

A. Characteristics of the Study Variables – Characteristics of the study variables will be evaluated to validate assumptions needed for the statistical test procedures. Histograms will be studied and tests performed to determine consistency of variables with basic assumptions. For continuous variables, the assumption of normality and equality of variance will be tested. For other variables, the numerical density of the observations will be evaluated to determine appropriateness of chosen categorical procedures. These evaluations include limitations of tabulated cell frequencies on the use of Chi-square tests.

B. Descriptive Statistics – Means, standard deviations, medians, and confidence intervals will be reported for all continuous variables. Dichotomous variables will be reported as percentages and the numerator and denominator will be reported and defined. Procedural, technical, and clinical success for the treatment arm will be reported. (See Figure 1, for definitions of procedural, technical and clinical success.) The frequency of major complications (as well as those of a specific type) will be reported as percentage rates with computed 95% two-sided exact confidence intervals (upper and lower).

III. Analyses of the Patient Populations – These analyses are intended to determine the similarity of two treatment groups and similarity of patients from different study sites with respect to important demographic or other variables, either known or suspected to have an influence on the outcome variables. The absence of similarity

for any variable will identify that variable as a potential covariate in subsequent safety and effectiveness analyses. Comparability analyses will be done with two-sided test with significance level 0.05.

A. **Comparability of Treatment Groups** – The demographic and prognostic variables measured at study entry will be compared between the EBV and control groups. Continuous variables will be compared with two-sample t-test or Wilcoxon rank sum test, and categorical variables will be compared with Fisher's exact test or Chi-square test.

B. **Study Site Comparability –** A set of important demographic or prognostic variables will be compared across study sites to determine homogeneity of study sites in patient characteristics. Factors found to differ significantly by study site will identify that variable or study site as a possible covariate in subsequent analyses. Smaller study sites with insufficient numbers of patients to allow a meaningful analysis will be combined into one or more pseudo-sites to allow the comparison to be done. The size of any pseudo-site created in this way will not exceed the size of the study site with the largest enrollment.

This analysis will not be used as a basis for data pooling across study sites. Data pooling will be done on a clinical basis, i.e., the sites used a common protocol, the sponsor adequately monitored the study to assure protocol compliance, and the data gathering and validation mechanisms were the same across all study sites (Meinert, 1986).

In addition, an analysis will be done on outcome to determine if there is a site by treatment interaction. Site by treatment interactions of a quantitative nature, i.e., all sites show the treatment to be beneficial, but perhaps to a different degree by study site, will not be considered to be an impediment to pooling. Site by treatment interactions that are qualitative in nature, i.e., the vast majority of sites show the treatment to be beneficial, but one or more sites show the treatment to be detrimental, will require extensive evaluation of the sites with contrary results to attempt to determine what factors at those sites led to the result (See expert statistical testimony from Dispute Resolution Panel transcript September 6, 2001).

In study sites with small numbers of patients, it will not be possible to evaluate a site by treatment interaction. The reason is that what may appear to be a site by treatment interaction may be a small numbers phenomenon. For example if there were only three patients from a given study site with one in the control group and two in the treated group, one success in each arm will appear to be a site by treatment interaction (100% success in the controls but 50% success in the treated arm). Obviously this example demonstrates the small numbers issue mentioned above.

IV. Analysis Populations

A. Geographic Cohort: VENT was conducted in U.S. and European sites. The PMA analysis primary cohort will be based on U.S. sites only. Section VIII below describes analysis of European data.

B. Primary: The primary endpoints, and all secondary endpoints, will be analyzed as an intent-to-treat (ITT) population defined as all randomized patients analyzed by the groups to which they were randomly assigned. Consenting patients who qualify for the study and are randomized into the study will be included in the ITT population, regardless of the actual treatment received. Consented patients 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 intent-to-treat group.

For the ITT analysis, missing data points will be imputed using a non-parametric multiple imputation in which patients withdrawn from the study will be randomly assigned outcomes by grouping on demographic and prognostic characteristics including treatment assignment maintaining masking, matching the characteristics to the withdrawn patients, and randomly selecting the result for the missed observation from the results for patients with similar characteristics by methods such as "hot deck" imputation or imputation by regression (Little and Rubin, 2002). The imputation will be stochastic imputation to preserve the variability of the imputed value. Also, the imputation will be done in a manner that is consistent with the assumptions of multiple imputation theory including missing at random to the extent possible. If the missing at random assumption is clearly violated by the data, other procedures including selection modeling and pattern mixture modeling will be considered.

C. Safety: The modified Intent-to-Treat (mITT) population is defined as all randomized patients who received study-directed treatment (EBV or control) and had any follow-up. The mITT population will be used for the safety analyses.

D. Secondary: Secondary analyses will be performed on a Completed Cases (CC) basis and Per-Protocol (PP) basis. The CC population is defined as all randomized and eligible patients who received study-directed treatment and had 6 months of follow-up. The PP population is defined as all randomized patients entirely consistent with the protocol and who received study-directed treatment and had 6 months of follow-up. Any visits where protocol violations occurred will be reviewed for possible exclusion from impacted analyses. These analyses will evaluate study endpoints sorting patients by the actual treatment received regardless of their randomization assignment (Completed Cases). Additionally, a sort by Completed Cases including only patients who were enrolled and treated in compliance with the protocol (Per Protocol) will be done.

V. Effectiveness Analysis

Effectiveness Analysis – There are two primary effectiveness variables, percent change from baseline at day 180 in FEV_1 AND 6MWT. The protocol indicates that

both primary endpoints need to be statistically significant, i.e., the endpoints are co-primary endpoints.

A. **Preliminary Analysis of Primary Effectiveness Endpoint**: The primary effectiveness endpoints will each be expressed as a mean percent change from baseline at day 180 with corresponding standard deviations. After imputation to account for missing values, the two ITT arms will be compared using an unpaired t-test or Mann-Whitney-Wilcoxon test if t-test assumptions are not met. All tests of superiority will be done with a one-sided test at a significance level of 0.025.

B. **In-Depth Analysis of Primary Effectiveness Endpoint:** Following the initial analysis for mean percent change in FEV_1 and 6MWT using an unpaired t-test or Mann-Whitney-Wilcoxon test, a multivariate analysis will be performed. Since the data are continuous, the analysis plan will use a general linear model (GLM) or mixed linear model procedure (PROC MIXED) to evaluate the effect of potential covariates. The purpose of these analyses will be to determine if specific subsets of the independent variables can be useful in predicting dependent variables.

A screening procedure similar to that described by Hosmer and Lemeshow (2000) will be used to reduce the set of variables eligible for inclusion in the model to reduce the possibility of over-specification of the model. In this screening procedure, a model will be fit to the relevant outcome variable that includes the potential covariate by the mixed model. Covariates that result in statistically significant main effects or statistically significant interactions with the outcome variables with a p-value of 0.20 (if the list is large, this could be reduced to 0.15) or less are allowed to enter the initial model. To be retained in the model, the p-value for any factor in the model must be 0.05 or less.

Potential covariates are shown in the tables below with required test methodology where applicable. Definitions of the potential covariates are listed in Appendix A. The High Resolution Computed Tomography Scan (HRCT) Core Lab prospectively assigns a target lobe to all patients prior to randomization. Most variables that are defined by HRCT analysis and target lobe will have a subset with both treated and control patients (*e.g.*: Upper Lobe treatments will be compared to Upper Lobe controls). For covariates of this type that are retained in the model, the subset analysis will be a test of the treatment arm subset vs. the control arm subset utilizing the methods described in section III (A) above. Subsets defined by valve characteristics will not have control groups (*e.g.:* Valve Version and Lobar Exclusion treatment subsets will not have a control subset). For these potential covariates without matching control groups, the comparative group will be the remaining treatment patients (*e.g.:* Lobar Excluded treatment subgroup vs. Non-lobar Excluded treatment subgroup).

The potential volume and spirometric covariates are listed in Table 1. These will be measured by Spirometry, Plethysmography, and the HRCT Core Lab. HRCT's are taken at the patient's residual volume (RV) and at total lung capacity (TLC). Some potential covariates require data from HRCT at RV, some from HRCT at TLC, and some require data from both.

Table 1: Plethysmography, Spirometry and HRCT Derived Baseline Potential Covariates*

Variable	Plethysmography /Spirometry	HRCT at TLC	HRCT at RV	TLC and RV HRCT †
RV	X			
RV % Predicted	X			
TLC	X			
TLC % Predicted	X			
RV/TLC	X			
VC	X			
FVC	X			
FVC % Predicted	X			
FEV_1	X			
FEV_1 % Predicted	X			
FEV_1/FVC	X			
DS of Target Lobe		X	X	
Ipsilateral DS Heterogeneity		X	X	
Thorax DS Heterogeneity		X	X	
For RUL Treatment Subset only: RUL-RML Heterogeneity		X	X	
Target Lobe Volume % of TLC		X		
Target Lobe Volume % of RV			X	
Total DS		X	X	
Max DS other than target		X	X	
Min DS other than target		X	X	
Target Lobe DS % Delta				X
Target Lobe Volume % Delta				X
Ipsilateral TLC to RV Absolute Volume Change Heterogeneity				X
Ipsilateral TLC to RV % Volume Change Heterogeneity				X
Fissure Score		X		

*Variables marked in multiple columns will be repeated by test. All tests are at Baseline.
†Some variables require information from both the TLC and RV HRCT

The Agency asked Emphasys to track the results of valve design changes made during the trial. This is listed in the Table 2 below along with other baseline and procedural potential covariates.

Table 2: Other Baseline and Procedure Potential Covariates

	Baseline	Procedure	Baseline V/Q
Gender	X		
Age	X		
Site	X		
6MWT	X		
Cycle Ergometry	X		
BMI	X		
Target Lobe		X	
RUL Treatment		X	
Upper vs. Lower		X	
Right vs. Left		X	
Valve Version*		X	
Large Valve*		X	
Small Valve*		X	
Valve Combos*		X	
Valve Expectorated*		X	

*These potential covariates apply only to the treatment arm and do not have a control arm subset.

The protocol states that the effect of atelectasis will be assessed. Atelectasis will be scored by the HRCT Core lab (see Table 3). A concern raised by the Agency during the IDE review process was the ability to detect valve migration and technical success. Based on this concern, the HRCT Core Lab will assess valve position and determine whether complete lobar exclusion was retained at day 180. This could be a potential covariate and is listed below in Table 3.

Table 3: 180 Day Follow-up HRCT Potential Covariates

Variable	180 Day HRCT at TLC	Baseline and 180 Day HRCT at TLC	Baseline and 180 Day HRCT at RV
Target Lobe Atelectasis Score		X	X
Lobar Exclusion (yes/no)*	X		

*These potential covariates apply only to the treatment arm and do not have a control arm subset.

Potential relevant covariate interactions will be also screened for inclusion in the multivariate model. These potential covariate interactions are listed below in Table 4.

Table 4: Composite or Other Potential Covariate Interactions

Composite or Interactions	Definition
Baseline BODE	BMI, FEV1% predicted, MMRC, 6MWT Composite Score
NETT Strata	Randomization Stratification by Baseline Cycle Ergometry and HRCT at TLC
NETT Strata by Right vs. Left	Randomization Stratification by Baseline Cycle Ergometry and HRCT at TLC, by Right or Left Lung
Lobar Exclusion by Fissure Interaction*	Fissure Score = 3 AND Lobar Exclusion = Yes vs. all others
Lobar Exclusion, Fissure, and RV % Predicted*	Lobar Exclusion = Yes AND Fissure Score = 3 AND RV % Predicted ≥ 200% vs. all others
Lobar Exclusion, Fissure, and RV % Predicted*	Lobar Exclusion = Yes AND Fissure Score = 3 AND RV % Predicted ≥ 200% vs. all others
Right vs. Left by TLC Heterogeneity	Right vs. Left Lung AND Ipsilateral DS Heterogeneity Max at TLC ≥ 10% or < 10%
Right vs. Left by RV Heterogeneity	Right vs. Left Lung AND Ipsilateral DS Heterogeneity Max at RV≥ 10% or < 10%
Upper vs. Lower by TLC Heterogeneity	Upper or Lower Lobe AND Ipsilateral DS Heterogeneity Max at TLC ≥ 10% or < 10%
Upper vs. Lower by RV Heterogeneity	Upper vs. Lower Lobe AND Ipsilateral DS Heterogeneity Max at RV ≥ 10% or < 10%
Large Valve by Right vs. Left*	Right vs. Left Lung AND Large Only vs. Others
Small Valve by Right vs. Left*	Right vs. Left Lung AND Small Only vs. Others
Expectorated Valve Replaced*	Valve(s) Expectorated AND Lobar Exclusion = Yes vs. all others
Expectorated Valve Not Replaced*	Valve(s) Expectorated AND Lobar Exclusion = No vs. all others
Lobar Exclusion by Site*	Technical Success by Site
First Case by Site*	First Treatment Case by Site vs. All Others
Learning Curve by Site*	Experience Level: Case 1, vs. 2 & 3, vs. 4-6, vs. 7-9, vs. ≥10 by site.

*These potential covariates apply only to the treatment arm and do not have a control arm subset.

Those variables found to be statistically significant via the mixed model analysis will be further analyzed in a follow-on sub-group analysis. In order to limit the impact of multiple hypothesis testing, univariate sub-set analysis will be performed using only those independent variables that remain in the mixed model. For those analyses performed on the remaining independent variables, appropriate adjustments will be made to account for impact on Type I error using methods described by Hochberg (1988).

C. **Secondary Effectiveness Analysis:** All secondary effectiveness analyses will be tests of superiority (*i.e.:* one-sided level 0.025 tests).

The following secondary endpoints (all at 180 days) will be tested in the same manner as the primary analysis of the primary effectiveness endpoints: mean % change in St. George's Respiratory Questionnaire, mean absolute change in mMRC, mean absolute change in room oxygen requirement (average liters per day), and mean absolute change in maximum workload as measured by cycle ergometry. Appropriate adjustments will be made to account for impact on Type I error using methods described by Hochberg (1988).

Each continuous endpoint will be expressed as a mean percent change from baseline or mean absolute change from baseline with corresponding standard deviation and the two arms will be compared initially using an unpaired t-test (or Mann-Whitney-Wilcoxon test if t-test assumptions are not met). Room oxygen requirement is calculated from the O_2 parameters reported by the patient (flow rates and hours at sleep, exertion and continuous) and will be reported as an average liters per day.

A multivariate analysis will be performed using a mixed model for the secondary effectiveness endpoints. The methodology and rationale is described in Sections V-B and V-F of this plan. The potential covariates specified in Section V-B will also be used in this analysis.

D. **Additional Analysis:** Additional analyses will evaluate the difference in the proportion of patients between treated groups that experience an improvement in FEV_1 of 15% or more. A second analysis will evaluate the proportion of patients in each group that achieve an improvement in the 6MWT of 15% or more. Analyses will also be performed on the mean % change in RV, mean % change in DLCO, mean absolute change in QWB, mean absolute change in BODE index, per-patient decrease in oxygen requirement (for patients who enter the study on supplemental oxygen) and per-patient increase in oxygen requirement (for all patients). These additional analyses will be considered exploratory and will not be used as the basis for labeling claims.

Because this a chronic, progressive disease, an analysis of an individual's 180 day endpoint compared to their baseline may understate the clinical utility. Improvement in FEV_1 will also be evaluated by testing the difference in the proportion of patients between treated groups that show no decline. (*i.e.:* proportion of patients whose 180 day % change in $FEV_1 \geq 0\%$). An identical analysis will be performed to evaluate 6MWT results (*i.e.:* proportion of patients whose 180 day % change in 6MWT $\geq 0\%$).

Technical, procedural, and clinical success for the treatment will be determined using the algorithm in Figure 1. Technical success is the rate of EBV cases in which lobar exclusion (complete exclusion of the target lobe, meaning that all airways feeding

that lobe are blocked with a valve) is achieved. Successful lobar exclusion will be determined by the HRCT core lab utilizing the 180 day HRCT at TLC. Procedural Successes are Technical Successes that do not reach the Major Complication Composite endpoint at 180 days. Clinical Successes are the Procedural Successes in which the primary clinical endpoint of 6MWT is 15% or greater over baseline at 180 days or in which daily oxygen consumption is decreased.

E. **Stratification Analysis:** Results for the primary safety and effectiveness endpoints will also be displayed for each of the four possible baseline stratification combinations (upper lobe/low exercise capacity, upper lobe/high exercise capacity, non-upper lobe/low exercise capacity, non-upper lobe/high exercise capacity). There is no expectation or requirement that treatment effectiveness is the same in the four possible stratification groups, nor is there a necessity to demonstrate statistical significance of the primary endpoints in all four subgroups.

F. **Further Analysis:** The percent receiving LVRS, lung transplant, or dying will be compared using a Fisher's exact test not controlling for study site and using logistic regression to control for study site. A Kaplan-Meier survival method will be used to compare the time to death from the time of randomization and an additional Kaplan-Meier analysis will be used to compare the time to a composite of lung transplant, LVRS and death from the time of randomization; a log rank test will be used to compare Kaplan-Meier curves between arms. The two Kaplan-Meier analyses will be followed by a multivariate Cox proportional hazards regression to adjust the analyses for possible covariates. The list of variables to be included in the screening and model building are included in Section V (B) above.

G. **Categorical-Variable, Multivariate Analysis:** The multivariate method for categorical variables is logistic regression.

Logistic regression is widely used and well understood method of risk analysis. Kleinbaum (1994) and Hosmer and Lemeshow (2000) provide an excellent discussion of the method and its properties. The model is based on

$$\ln [p/(1-p)] = \alpha + \Sigma\beta_i X_i$$

where ln is the natural logarithm, p is the probability of outcome (success or adverse event), α is the intercept of the fit line, β_i is the coefficient corresponding to the ith variable known or suspected of being predictive of outcome, and X_i is the observed value of that variable. The X_i can be either continuous or ordered categorical variables. The model described by the equation above can also consider separate terms that measure the combination of variables known or suspected of being predictive of outcome. These combinations, known as interactions, provide evidence of subtle relationships between the predictive variables that may increase or decrease the risk of outcome when the patient exhibits certain combinations of risk factors. Any term remaining in the final model is associated with the outcome

in a statistically significant way, and its odds ratio and P-value are adjusted for the presence of other factors remaining in the model.

The screening procedure described in Section V (B) will be used to reduce the set of potential covariates eligible for inclusion in the model to reduce the possibility of over-specification of the model.

Those variables found to be statistically significant via the logistic regression analysis will be further analyzed in a follow-on sub-group analysis. In order to limit the impact of multiple hypothesis testing, univariate sub-set analysis will be performed using only those independent variables that remain in the multivariate model. For those analyses performed on the remaining independent variables, appropriate adjustments will be made to account for impact on Type I error using methods described by Hochberg (1988).

VI. Safety Analyses

A. The proportion of patients treated with the EBV who reach the Major Complication Composite (MCC) will be reported along with exact 95% confidence intervals. The MCC endpoint is met when one of the composite complications occurs. There can only be one complication composite endpoint per patient; however, each event will be recorded and reported along with other complications. There is no expectation or requirement that the MCC rate is the same in both arms.

In addition, a logistic regression will be done to determine the relationship between potential covariates and the MCC. The screening procedure described in Section V (B) will be used to reduce the set of potential covariates eligible for inclusion in the model to reduce the possibility of over-specification of the model.

Those variables found to be statistically significant via the logistic regression analysis will be further analyzed in a follow-on sub-group analysis. In order to limit the impact of multiple hypothesis testing, univariate sub-set analysis will be performed using only those independent variables that remain in the multivariate model. For those analyses performed on the remaining independent variables, appropriate adjustments will be made to account for impact on Type I error. These adjustments will be made using methods described by Hochberg (1988).

B. Secondary safety analyses will be done comparing the proportion of patients by treatment that experiences any complication. The rates and 95% exact confidence intervals will be presented.

Complications will be categorized into clinically relevant groups (*e.g.:* stable pneumothorax with no intervention, pneumothorax resolved with chest tube insertion in less than 7 days, prolonged air leak, etc.). Complications occurring in the treatment arm will be further categorized as Device-related, Procedure-related, or neither and by those occurring during procedure hospitalization and those occurring post-discharge.

Rehospitalization rates will be reported by treatment arm on a Per-Patient (PPt) basis and on a Per-Event (PE) basis. The Per-Patient rehospitalization rate is the proportion of patients who were re-admitted post-discharge. An individual patient will only be counted once in the PP no matter how many times they are readmitted during the follow-up period. The Per-Event rehospitalization rate is the proportion of hospital readmissions per treatment arm including multiple readmissions per individual patient.

C. Early and Late Adverse Events: In order to assess the safety profile of the valve implant without the interference of the procedure effect, a subset analysis will be performed comparing Early Adverse Events (0-90 days post procedure) vs. Late Adverse Events (91-180 days post randomization). For those patients with 365 day follow-up, a subset analysis will also be performed comparing Adverse Events that occur from day 181-365 post randomization.

VII. Patient Accountability and Missing Data

The number and proportion of patients eligible for and compliant with each follow-up examination will be presented. This patient accountability data will be presented by follow-up time and by treatment group. Patients who withdraw from the study will be tabulated with the reasons for the withdrawal. The withdrawals will be tabulated by treatment group. If the proportion of patients withdrawn is larger than the 15% from either arm, an analysis of the demographic and prognostic characteristics will be made between patients who withdraw and those who remain in the study. For continuous variables, parametric or non-parametric analysis of variance will be used. For categorical variables, Chi-square or Fisher's exact test will be applied. Patients eligible for follow-up visits will include patients remaining in the trial. Patients that die or withdraw are not eligible for follow-up. The completeness of follow-up, i.e., the proportion seen at any follow-up visit will be computed in two ways: it will be computed based on the eligible patients and it will be computed based on eligible patients minus patients lost to follow-up.

The evaluation of withdrawn patients presents a special concern. All clinical studies analyze the results based on the evaluable patients, i.e., those who complete the study. Because withdrawn patients do not have final data, they present a problem. The statistical community (Pocock, 1983, Meinert, 1986, Friedman et al., 1985, and ICH Guidance, 1998) recommends that multiple sensitivity analyses be conducted to determine the robustness of the result in patients who complete the study. The intention of these analyses is to demonstrate that the results obtained from the evaluable patients are not biased.

As a result, sensitivity analyses using several methods including multiple imputation analyses will be conducted to evaluate the robustness of the study result accounting for missing observations. One imputation which is biased against an effective test device will randomly assign patients outcome using random sampling of the distribution of the control group as the missing value for any study patient. Often the Food and Drug Administration requires sponsors to impute the worst outcome for patients who do not complete the trial and this imputation will be

presented in this trial as one of several sensitivity methods. The worst outcome for this analysis will be the worst performer for each primary endpoint from the respective arm for patients that completed the trial.

VIII. Analysis of Pooled U.S. and European Data

Emphasys Medical had submitted an IDE supplement dated January 8, 2004 stating our intention to enroll subjects at both U.S. and European sites. Based on recommendations from the Agency at the October 6, 2006 pre-PMA meeting, this plan has been amended to include only the U.S. data as the primary analysis cohort or pivotal data cohort. Emphasys Medical will provide an abbreviated data set for the European cohort. This abbreviated analysis will include safety summary analyses of the Major Complication Composite and other complications. Additionally, Emphasys will report descriptive statistics for the primary effectiveness endpoints using the Completed Cases population.

IX. Analysis of EBV-OTW data

Emphasys Medical submitted an IDE supplement and has received approval to introduce a second generation EBV into the VENT study. As a condition of this approval, Emphasys agreed not to "pool" the data associated with the first generation "EBV-OTW" product with that of the second generation "Transcopic (TS)" product. Given this stipulation, the company will randomize a "new" set of a minimum of 246 subjects (164 treatment and 82 control subjects) using the "TS" product. It is intended that this cohort of 246 minimum "TS" patients will serve as the "pivotal trial" data set.

Emphasys does not intend to pursue approval for the "EBV-OTW" product configuration. As such the company intends to follow those patients treated in the "EBV-OTW" phase of the study for one year for safety observation per the original protocol. It is proposed that the control group from this phase of the study be given the opportunity to be re-randomized (assuming they continue to meet inclusion/exclusion criteria) into the "TS" phase of the study once they have passed the 180 day follow-up visit. The "EBV-OTW" cohort includes 46 treatment and 19 control subjects.

Given that the company does not plan to pursue approval for the "EBV-OTW" product configuration, an abbreviated statistical analysis for this cohort will be performed. This abbreviated analysis will include summary analyses of the Major Complication Composite and other complications. Additionally, Emphasys will report descriptive statistics for the primary effectiveness endpoints using the Completed Cases population. The company does not expect to reach statistically significant endpoints in this cohort and would anticipate that the outcome (positive or negative) will not be used as the basis for PMA approval, labeling claims, or labeling restrictions for the TS products.

X. Malfunction Analyses

Results of any device malfunctions and their sequelae are to be presented descriptively. The rate and exact 95% confidence intervals will be computed.

XI. Statistical Software

The parametric and non-parametric analysis of variance and other primary analyses will be done using SAS, Version 8.2 for Personal Computers or later. The Chi-square and Fisher's exact tests will be done with StatXact, Version 4.01 for Windows or later. Some descriptive analyses may be done with SYSTAT 10.0.

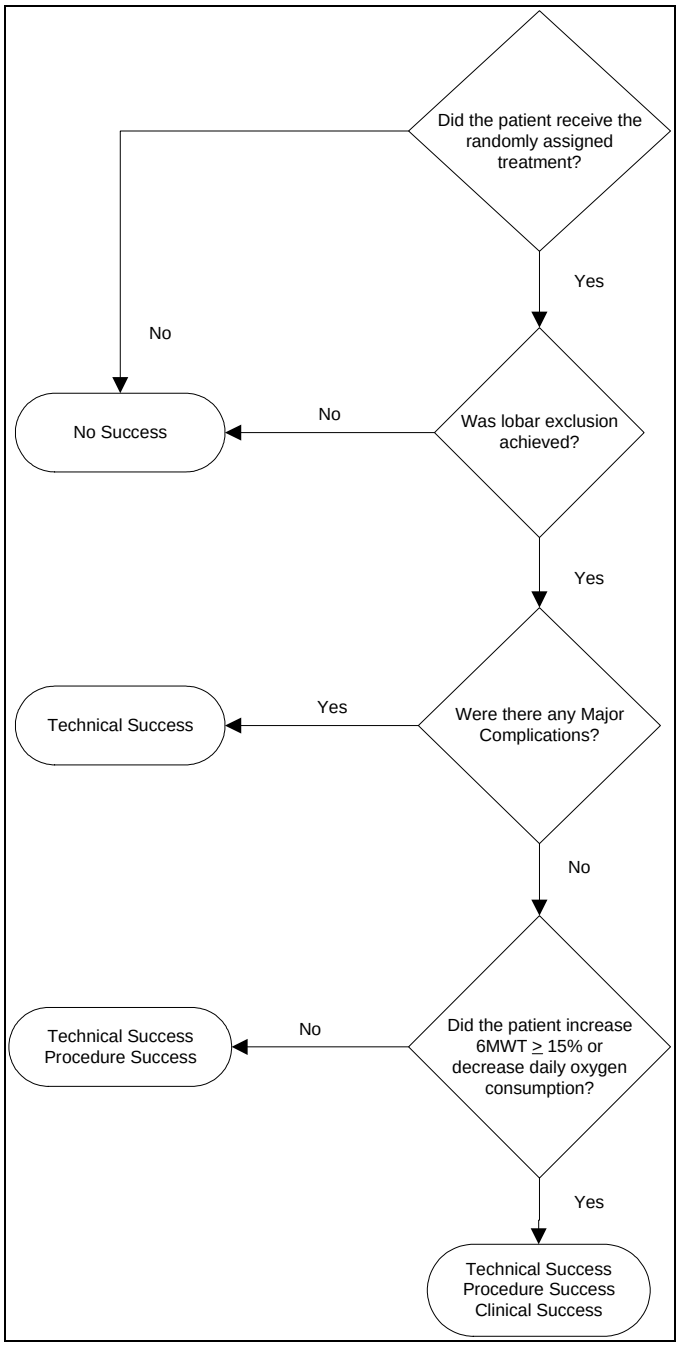

Figure 1: Treatment Arm Technical, Procedural, and Clinical Success Criteria

References:

1. Meinert, C. (1986). *Clinical Trials: Design, Conduct, and Analysis*. Oxford University Press, New York.

2. Medical Devices Dispute Resolution Panel Meeting of September 6, 2001 (Panel Transcript)

3. Kleinbaum, D. (1994). *Logistic Regression*. Springer-Verlag, New York.

4. Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. John Wiley and Sons, New York.

5. Pocock, S. (1983). *Clinical Trials (A Practical Approach)*. John Wiley and Sons, Chichester.

6. Friedman, L., C. Furberg, and D. DeMets. (1985). *Fundamentals of Clinical Trials*. Mosby Year Book, St. Louis.

7. International Conference on Harmonisation. Guidance on Statistical Principles for Clinical Trials. Availability. Federal Register 63 (179): 49583-49598, September 16, 1998.

8. Little, R. and D. Rubin. (2002). *Statistical Analysis with Missing Data*. John Wiley and Sons, New York.

9. Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75:800-802.

Supplement A: Definitions

Variable	Definition
RV	Residual Volume by Spirometry and Plethysmography (TLC – Max (FVC or VC))
TLC	Total Lung Capacity by Plethysmography
VC	Vital Capacity by Plethysmography
FVC	Forced Vital Capacity by Spirometry
FEV_1	Forced Expiratory Volume in the First Second by Spirometry
Target Lobe	RUL, RLL, LUL, or LLL
RUL	RUL Target vs. Others
DS	Density Score by Lobe
Total DS	Sum of all Lobar Density Scores (including RML)
Ipsilateral DS Heterogeneity	Target – Max DS of Non-treated Lobes within Treated Lung [e.g.: RUL – Max(RML, RLL)]
Thorax DS Heterogeneity	Target DS – Max DS of All Non-treated Lobes [e.g.: RUL - MAX (RML, RLL, LUL, LLL)]
Target Lobe DS % Delta	(TLC HRCT – RV HRCT)/TLC HRCT Density Score of Target Lobe
Target Lobe Volume % Delta	(TLC HRCT Volume – RV HRCT)/TLC HRCT Volume of Target Lobe
Ipsilateral TLC to RV Absolute Volume Change Heterogeneity	(Target Lung TLC Volume – RV Volume) - (Ipsilateral non-Target Lung TLC – RV Volume) [e.g.: (RUL Volume TLC RV Volume Delta) - ((RML+RLL TLC Volume)-(RML+RLL RV Volume))]
Ipsilateral TLC to RV % Volume Change Heterogeneity	(Target Lobe TLC Volume – RV Volume)/Target Lobe TLC Volume - (Ipsilateral non-Target Lobe TLC – RV Volume)/non-Target TLC Vome [e.g.: (RUL TLC-RV Volume Delta)/RUL - ((RML+RLL TLC Volume)-(RML+RLL RV Volume))]
Fissure Score	Categorical Fissure Assessment Surrounding Target Lobe 1 = Absent 2 = Incomplete 3 = Complete
NETT Strata	1 = Upper-Lobe Predominance, Low Baseline Exercise Capacity* 2 = Upper-Lobe Predominance, High Baseline Exercise Capacity 3 = Non-Upper-Lobe Predominance, Low Baseline Exercise Capacity 4 = Non-Upper-Lobe Predominance, High Baseline Exercise Capacity
Target Lobe Atelectasis Scores	% change in Target Lobe Volume at TLC and at RV between Baseline and 180-day HRCT
Lobar Exclusion	Yes, If all bronchial pathways to target lobe are sealed by valve(s) - adjudicated by HRCT Core Lab
Valve Combo	1. Ver1 Small ONLY 2. Ver1 Small AND Large 3. Ver2 Small ONLY 4. Ver2 Small AND Large 5. Large Only
Valve Version	Any procedure with a Version 1 Valve (Valve Combo 1 or 2) vs. all others
Large Valve	Any procedure with Only Large Valves (Valve Combo 5) vs. all others
Small Valve	Any procedure with Only Small Valves (Valve Combo 1 or 3) vs. all others

*Low Exercise Capacity by Cycle Ergometry $\leq$ 25W for Women, $\leq$ 40W for Men