STATISTICAL ANALYSIS PLAN

Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation

(The COAPT Trial)

Protocol # 11-512

Version 8.3

June 4, 2018

Contents

1 Introduction

This document contains the Statistical Analysis Plan (SAP) for the COAPT Trial, protocol 11-512. This plan is based on the Version 8.0, October 12, 2016 study protocol.

2 Purpose

The objective of the COAPT Trial is to evaluate the safety and effectiveness of the MitraClip System for the treatment of moderate-to-severe (3+) or severe (4+) functional mitral regurgitation (FMR) in symptomatic heart failure (HF) subjects who are treated per standard of care and who have been determined by the site's local heart team as not appropriate for mitral valve (MV) surgery. Eligible subjects will be randomized in a 1:1 ratio to the MitraClip device (Device group) or to no MitraClip device (Control group).

As these patients typically do not undergo surgery, the appropriate control group is non-surgical treatment for management of symptoms from MR. The trial has two co-primary objectives, including one primary safety objective and one primary effectiveness objective. The trial also has two secondary safety objectives and several secondary effectiveness objectives.

2.1 Primary Objectives

The primary safety objective is to demonstrate that freedom from a composite of device-related complications in the Device group at 12 months is greater than a performance goal of 88%. The safety composite includes Single Leaflet Device Attachment (SLDA), device embolizations, endocarditis requiring surgery, Echocardiography Core Laboratory confirmed mitral stenosis requiring surgery, LVAD implant, heart transplant, or any device related complications requiring non-elective cardiovascular surgery.

The primary effectiveness objective is to demonstrate that subjects in the Device group experience a lower rate of recurrent HF hospitalization than subjects in the Control group.

2.2 Secondary Objectives

There are two secondary safety objectives:

- To demonstrate that freedom from death (all-cause), stroke, myocardial infarction (MI), and non-elective (i.e., urgent or emergent) cardiovascular surgery for device related complications at 30 days post-procedure in the Device group is greater than 80%

- To demonstrate that the hazard ratio of all-cause mortality at 12 months between the Device and Control groups is less than 1.5

There are several secondary effectiveness objectives as described in the COAPT Protocol Version 8.0 Section 3.2.

3 Study Design

The COAPT Trial is a prospective, randomized, parallel-controlled, multicenter clinical evaluation of the MitraClip device for the treatment of clinically significant functional mitral regurgitation in symptomatic HF subjects who are treated per standard of care and who have been determined by the site's local heart team as not appropriate for MV surgery. Eligible subjects will be randomized in a 1:1 ratio to the MitraClip device (Device group) or to no MitraClip device (Control group).

Investigational sites will attempt to recruit consecutive subjects who meet trial eligibility criteria. Approximate 610 subjects will be randomized at up to 100 investigational sites. Investigators at any site may not register more than 15% of the total of 610 randomized subjects. There is no minimum number of subjects to be registered at any site. Up to an additional 150 roll-in subjects, up to 3 per site, may be treated. Once the site completes the roll-in phase, the randomization phase will begin. Once the randomization phase has begun, all subsequent subjects must be randomized into the trial. Roll-in subjects will not count toward the target randomization of 610 subjects.

Upon randomization, subjects will be scheduled for a "Treatment" visit. At the "Treatment" visit (between 1 and 14 days after randomization), Device group subjects will undergo the MitraClip procedure. Control group subjects will be seen by the HF specialist investigator at the "Treatment" visit (between 1 and 14 days after randomization), and will undergo a physical exam, including vital signs, cardiac health status and evaluation of HF medications. Thus, scheduled visits will be performed consistently between the two groups to ensure comparable levels of contact between the two groups.

Subjects will be followed at 1 week (phone contact), 30 days, 6 months, 12 months, 18 months, 24 months, and annually thereafter through 5 years from the "Treatment" visit.

Control group subjects will not be allowed to undergo the MitraClip procedure until completion of the 24-month visit. MitraClip intervention in the Control group before the 24-month visit will be considered as a protocol deviation. Mitral valve surgery in either group will be considered a protocol deviation, unless the subject experiences a complication (e.g., endocarditis, clip detachment or leaflet injury).

4 Study Primary and Secondary Endpoints

4.1 Primary Safety Endpoint

The primary safety endpoint is a composite of Single Leaflet Device Attachment (SLDA), device embolizations, endocarditis requiring surgery, Echocardiography Core Laboratory confirmed mitral stenosis requiring surgery, LVAD implant, heart transplant, or any device related complications requiring non-elective cardiovascular surgery at 12 months.

4.2 Primary Effectiveness Endpoint

The primary effectiveness endpoint is recurrent HF hospitalizations through 24 months (analyzed when the last subject completes 12 months of follow-up). For subjects who undergo MV surgery, LVAD implant or heart transplant during the follow-up, the hospitalizations for the MV surgery, LVAD or heart transplant intervention will be treated as HF hospitalizations in the analysis. For subjects in the Control group who receive the MitraClip device due to heart failure or cardiac symptoms, the hospitalizations for the MitraClip procedure will be treated as HF hospitalization in the analysis.

4.3 Secondary Safety Endpoints

The secondary safety endpoints are:

- A composite of all-cause death, stroke, MI, or non-elective cardiovascular surgery for device related complications in the Device group at 30 days
- All-cause mortality at 12 months

4.4 Secondary Effectiveness Endpoints

The secondary effectiveness endpoints are:

- MR severity at 12 months
- Hierarchy of death and recurrent HF hospitalization (analyzed when the last subject completes 12 months of follow-up)
- Change in quality of life as measured by the Kansas City Cardiomyopathy Questionnaire (KCCQ) at 12 months from baseline
- Change in Six-Minute Walk test distance (6MWD) at 12 months from baseline
- NYHA Functional Class of I/II at 12 months
- Change in LVEDV at 12 months from baseline
- Recurrent hospitalizations - all-cause (analyzed when the last subject completes 12 months of follow-up)

5 Statistical Methods

5.1 Randomization and Blinding

Subjects will be randomized in a 1:1 ratio to the Device or Control groups. Randomization will be stratified by site and cardiomyopathy etiology (ischemic or non-ischemic). Randomization will be performed using permuted blocks (random block sizes) and block

sizes will not be revealed to sites. The Electronic Data Capture system will be used to randomize subjects.

The subject, some site personnel and some Sponsor personnel will be aware of treatment assignment. Sponsor statisticians will not have access to any data that combines outcomes with treatment assignment prior to performing the final analysis. No Sponsor personnel will have access to outcome data summarized by treatment until the primary endpoint analysis is complete. All events constituting primary or secondary endpoints will be adjudicated by the independent Clinical Events Committee (except embolization, which will be site reported, and MV Stenosis and SLDA, which will be assessed by the Echocardiography Core Lab), along with relationship to the MitraClip device or procedure. Every effort will be made to blind the Clinical Events Committee (CEC) to the subject's treatment assignment.

5.2 Data Monitoring Committee

A Data Monitoring Committee (DMC) will meet periodically at a frequency set out in the DMC charter to monitor the safety of subjects enrolled in the study. The DMC may recommend terminating the study early if there are safety concerns.

5.3 Demographic and Baseline Characteristics

The Device and Control groups will be compared with respect to demographic and baseline characteristics. The two groups are expected to be well balanced at baseline due to randomization. Comparisons will be made with respect to age, sex, race and ethnicity, cardiovascular conditions, diabetes, renal disease, prior MI, prior stroke, NYHA Class, previous cardiac surgery, prior HF hospitalization, LVEF, left ventricular size, creatinine levels, STS mortality risk score, MR severity, KCCQ score, 6-minute walk distance and baseline medications. Continuous variables will be analyzed using the t-test or using non-parametric methods and categorical variables will be compared using Fisher's Exact test.

5.4 Analysis Populations

The Intention to Treat, As Treated, Per Protocol and Safety Analysis populations are defined below.

5.4.1 Intention to Treat (ITT) Population

The Intention to Treat population will consist of all subjects randomized in the trial. All subjects will be analyzed in the group randomized, regardless of the treatment actually received.

5.4.2 As Treated (AT) Population

The As Treated population is defined as consisting of all randomized subjects according to the treatment they received. Subjects who experience a death or HF hospitalization prior to a MitraClip procedure will be considered to be in the Control group regardless of their initial

randomization. Subjects who experience a death or HF hospitalization after (but not prior to) a MitraClip procedure will be considered to be in the MitraClip group regardless of their initial randomization. For patients who do not experience a death or HF hospitalization at any time during follow-up, they will be assigned to the group that constituted > 50% of their follow-up duration. For example, if an event-free subject with 24-month follow-up received a MitraClip at 9 months, they would be assigned to the MitraClip group, whereas if an event-free subject with 24-month follow-up received a MitraClip at 15 months, they would be assigned to the Control group.

5.4.3 Per Protocol (PP) Population

The Per Protocol population is defined as consisting of all randomized subjects who meet all major study inclusion criteria and none of the major exclusion criteria for the trial, are treated according to the randomized assignment, and followed consistent with all major study processes. Subjects who do not complete the "Treatment" visit will be excluded from the PP population. Subjects who are randomized to the Device group but do not have a MitraClip procedure attempted between 1-14 days of the randomization will be excluded from the PP population. For subjects in the Control group who receive the MitraClip device, follow-up data after the date of the procedure will be excluded from analysis. For subjects in either group who undergo other major intervention for HF (e.g. MV surgery, LVAD implant or heart transplant), follow-up data after the date of the intervention will be excluded from analysis.

5.4.4 Safety Analysis (SA) Population

The Safety Analysis population is defined to consist of randomized Device group subjects in whom a MitraClip procedure is attempted.

5.5 Analysis of Safety

5.5.1 Primary Safety Endpoint

A composite of device related complications including Single Leaflet Device Attachment (SLDA), device embolizations, endocarditis requiring surgery, Echocardiography Core Laboratory confirmed mitral stenosis requiring surgery, LVAD implant, heart transplant, or any device related complications requiring non-elective cardiovascular surgery at 12 months will be the primary measure of safety.

All events constituting the primary safety endpoint will be adjudicated by the independent Clinical Events Committee (except embolization, which will be site reported, and Mitral Valve Stenosis and SLDA, which will be assessed by the Echocardiography Core Lab). SLDA is defined as unilateral MitraClip detachment from one leaflet, as assessed by the Echocardiography Core Laboratory as definite. In the case of MV surgery, SLDA can also be identified by operating physician.

The analysis of the primary safety endpoint is a test of the proportion of subjects free from the composite of safety events in the Device group at 12 months to a pre-specified performance goal of 88%.

Hypothesis:

The null and alternative hypotheses may be stated as:

$\mathbf{H_0}$: $P_D(12) \leq 0.88$ vs.

$\mathbf{H_1}$: $P_D(12) > 0.88$

where, $P_D(12)$ is the proportion of subjects free from safety events at 12 months in the Device group. This endpoint will be evaluated at the one-sided significance level of 5%.

Since this is a 12-month endpoint, all data will be truncated at 12 months for the analysis. The Kaplan Meier survival estimate, together with variance estimated by the Greenwood method, will be used to set up the test of the null hypothesis as a Z-test. The null hypothesis will be rejected at the 5% level of significance if the test statistic is greater than 1.645.

The test statistic is:

$$\frac{\hat{P}_D - 0.88}{\sqrt{\hat{V}\left(\hat{P}_D\right)}}$$

where $\hat{P}_D$ is the Kaplan Meier survival estimate in the Device group. $\hat{V}\left(\hat{P}_D\right)$ is the Greenwood estimate of the variance.

The primary safety endpoint will be considered met if the null hypothesis is rejected at the 5% level of significance in the primary analysis described below. Sensitivity analyses will also be performed in addition to the primary analysis as detailed below.

5.5.1.1 Primary Analysis

The primary analysis will be performed in the Safety Analysis population. Time to event or time to censoring will be calculated from the procedure ("Treatment" visit) date. In subjects with events within 12 months from the "Treatment" visit, time to event will be calculated as the length of time from the date of the "Treatment" visit to the date of the first event. Subjects who died within 12 months without any event will be censored on the date of death. Subjects who withdraw within 12 months without any event will be censored on the date of withdrawal. Subjects who are not withdrawn, and who have not experienced a primary safety event through last follow-up with evaluable echocardiograms will be censored on the date that follow-up occurred.

5.5.1.2 As Treated (AT) Analysis

Data from all Device group subjects in the AT population will be analyzed similar to the primary analysis. For subjects who undergo other intervention for HF (e.g. MV surgery, LVAD implant or heart transplant), their follow-up information after the first intervention for HF will be excluded from the AT analysis if they have not experienced safety events prior to the date of the intervention.

5.5.1.3 Per Protocol (PP) Analysis

Data from all Device group subjects in the PP population will be analyzed similar to the primary analysis. For subjects who undergo other intervention for HF (e.g. MV surgery, CRT implant), their follow-up information after the first intervention for HF will be excluded from the PP analysis if they have not experienced safety events prior to the date of the intervention.

5.5.1.4 Tipping-point Analysis

Tipping point analysis on the primary safety endpoint at 12 months will be conducted as a sensitivity analysis. Subjects in the Safety Analysis population who withdraw from the study before completing a year of follow-up will be analyzed one at a time as having experienced a primary safety endpoint event on the date of withdrawal. In addition, patients with unevaluable echocardiograms at last follow-up will be analyzed one at a time as having experienced a primary safety endpoint event on the date that follow-up occurred.

5.5.1.5 Subgroup Analyses for Primary Safety Endpoint

5.5.1.5.1 Sex (male vs. female)

Subgroup analyses of sex will be performed. Major baseline and clinical characteristics will be summarized by sex.

A Kaplan-Meier analysis with stratification for sex will be performed for time to first event. Freedom from the safety composite at 12 months will be reported for the strata. An additional analysis will also be performed utilizing a Cox regression model for analyzing time to first event. The test for the sex effect will be performed at the 5% level of significance.

5.5.1.5.2 Etiology of Cardiomyopathy (ischemic vs. non-ischemic)

Subgroup analyses of cardiomyopathy etiology will be performed. Major baseline characteristics will be summarized by etiology of cardiomyopathy.

A Kaplan-Meier analysis with stratification for cardiomyopathy etiology will be performed for time to first event. Freedom from the safety composite at 12 months will be reported for the strata. An additional analysis will also be performed utilizing a Cox regression model for analyzing time to first event. The test for the effect of etiology of cardiomyopathy will be performed at the 5% level of significance.

5.5.1.5.3 Left Ventricular Ejection Fraction (LVEF > 40% vs. LVEF ≤ 40%)

Subgroup analyses of Central Echocardiography Core Laboratory assessed baseline LVEF will be performed. Major baseline and clinical characteristics will be summarized by baseline LVEF.

A Kaplan-Meier analysis with stratification for LVEF will be performed for time to first event. Freedom from the safety composite at 12 months will be reported for the strata. An additional analysis will also be performed utilizing a Cox regression model for analyzing time to first event. The test for the LVEF effect will be performed at the 5% level of significance.

5.5.1.5.4 Extreme Surgical Risk Status

Subgroup analyses of extreme surgical risk status (as assessed by the Central Eligibility Committee) will be performed. Major baseline and clinical characteristics will be summarized by extreme surgical risk status.

A Kaplan-Meier analysis with stratification for extreme surgical risk status will be performed for time to first event. Freedom from the safety composite at 12 months will be reported for the strata. An additional analysis will also be performed utilizing a Cox regression model for analyzing time to first event. The test for the effect of extreme surgical risk status will be performed at the 5% level of significance.

5.5.1.6 Analysis of Poolability of Study Sites

Analysis of the site effect will be performed. A Cox regression model will be utilized for analyzing the time to first event. The model will include the effects of site. The site effect will be tested at the 15% level of significance.

The study will involve up to 100 sites. It is likely that there will be wide variation in the number of subjects enrolled across sites (due to differences in site start-up times and differences in subject volumes enrolled at sites). If model convergence is not achieved, sites will be categorized into the categories described below. The model will then include the effects of site category. In addition, a descriptive summary of events will be provided by site.

- Large: Sites with more than 10 subjects

- Medium Large: Sites that enroll between 6 and 10 subjects
- Medium: Sites that enroll between 3 and 5 subjects
- Small: Sites that enroll fewer than 3 subjects

5.5.2 Secondary Safety Endpoints

5.5.2.1 Composite 30-Day Secondary Safety Endpoint

A composite of all-cause death, stroke, MI, or non-elective (urgent or emergent) cardiovascular surgery for device related complications in the Safety Analysis Population at 30 days will be used as a secondary measure of safety. The analysis of the secondary safety endpoint is a one-group test against a performance goal for the proportion of subjects in the Safety Analysis Population free from the composite of secondary safety events at 30 days. The null and alternative hypotheses may be stated as:

$\mathbf{H_0}$: $P_D(30) \leq 0.80$ vs.

$\mathbf{H_1}$: $P_D(30) > 0.80$

where, $P_D(30)$ is the proportion of subjects free from composite of secondary safety events at 30 days in the Safety Analysis Population. This endpoint will be evaluated at the one-sided significance level of 5% with an exact test for a single proportion. The denominator of the proportion will exclude subjects in the Safety Analysis Population who terminated from the study prior to the lower window of the 30-Day visit (Day 27) without any site reported adverse events.

Components of the secondary composite safety endpoint will be summarized and reported for the Device group.

5.5.2.2 All-Cause Mortality at 12 Months

The hazard ratio of all-cause mortality from the time of randomization to 12 months between the Device and Control groups is a secondary measure of safety. The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $HR \geq 1.5$ vs.

$\mathbf{H_1}$: $HR < 1.5$

where, HR is the hazard ratio of all-cause mortality at 12 months between the Device and Control groups.

Follow-up will be truncated at 12 months. A Cox regression model with only treatment effect as covariate will be used to estimate the hazard ratio. The one-sided 95% upper confidence bound for the HR will be calculated as HR + Z_α × SE estimation from the Cox regression model, where $Z_\alpha = 100(1-\alpha)$th percentile of the standard normal distribution. The null hypothesis will be rejected at the one-sided significance level of 5% if the one-sided 95% upper confidence bound for HR is less than 1.5.

If non-inferiority of Device group to Control group is demonstrated, the estimated HR is lower than 1, and the difference in the survival curves between Device group and Control group is maintained or increases after 12 months, then a superiority test of all-cause mortality will be performed in the pre-specified fixed sequential testing for all secondary effectiveness endpoints (See **Section 5.9 Multiplicity Adjustments**). All available follow-up through 24 months at the time of data cut-off will be included in the analysis. The null and alternative hypotheses for superiority test are stated as:

H_0: HR = 1.0 vs.

H_1: HR ≠ 1.0

The superiority null hypothesis will be rejected at the two-sided significance level of 5% if the 95% upper confidence bound for HR is less than 1.

5.5.2.2.1 Intention to Treat Analysis (Primary Analysis)

This analysis will be performed in the ITT population. Duration of follow-up will be calculated from the date of randomization.

In addition, analyses of this secondary safety endpoint will be performed on the AT and PP populations.

5.5.2.2.2 Landmark Survival Analysis

A landmark analysis for all-cause mortality will be performed, with follow-up starting at 30 days after randomization and through 12 months after randomization. This analysis will be identical to the primary analysis except that only subjects who have not been lost to follow-up, died or withdrawn by 30 days post randomization will be included and time to death or censoring will be calculated from 30 days.

5.6 Analysis of Effectiveness

5.6.1 Primary Effectiveness Endpoints

Treatment with the MitraClip device is expected to reduce the risk of recurrent HF hospitalization. Recurrent HF hospitalization is the primary effectiveness endpoint for the trial. Hospitalizations that are adjudicated by the CEC as related to HF using the pre-specified protocol definition will be included as events in the analysis. The primary effectiveness endpoint will be analyzed from the time of randomization until the last subject completes 12 months of follow-up. All follow-up data through 24 months at the time of data cut-off will be used.

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: RRR ≤ 0 vs.

$\mathbf{H_1}$: RRR > 0

where RRR is the relative risk reduction in the rate of recurrent HF hospitalization due to treatment with the MitraClip device. The Joint Frailty Model[1] (JFM) method will be used to model the recurrent hospitalization data and estimate the RRR.

RRR = (1- Hazard Ratio of Device vs. Control) = $(1 - e^{\beta})$ where β is the regression coefficient of the recurrent HF hospitalization in the JFM model.

The primary analysis in the ITT population and other analyses are listed below. The endpoint will be met if the null hypothesis in the primary analysis is rejected at the one-sided significance level of 5%.

5.6.1.1 Intention to Treat Analysis (Primary Analysis)

This analysis will be performed in the ITT population. Subjects who do not experience any HF hospitalizations and are alive will be censored on their data cut-off date. Subjects who die will be censored on the date of death. Subjects who withdraw from the study without experiencing a HF hospitalization will be censored on the date of withdrawal. Duration of follow-up will be calculated from the date of randomization. For a sensitivity analysis, duration of follow-up will also be calculated from the date of the "Treatment" visit or the randomization date if the "Treatment" visit is missing.

5.6.1.2 As Treated Analysis

The analysis in the AT population will be identical to the ITT analysis with the exception that it will be performed in the AT population. Duration of follow-up will be calculated from the date of the randomization. For Subjects who undergo other intervention for HF (e.g. MV surgery, LVAD implant or heart transplant) in either group, their follow-up information after the first intervention for HF will be excluded from the AT analysis (regardless of treatment group).

For a sensitivity analysis, duration of follow-up will also be calculated from the date of the "Treatment" visit or the randomization date if the "Treatment" visit is missing.

5.6.1.3 Per Protocol Analysis

The analysis in the PP population will be identical to the ITT analysis with the exception that it will be performed in the PP population. Duration of follow-up will be calculated from the date of randomization. For subjects who undergo other intervention for HF (eg. MV surgery, LVAD implant or heart transplant) in either group, their follow-up information after the first intervention for HF will be excluded from the PP analysis. For subjects in the Control group who receive the MitraClip device, their follow-up information after the MitraClip procedure will be excluded from the PP analysis (regardless of treatment group).

For a sensitivity analysis, duration of follow-up will also be calculated from the date of the "Treatment" visit.

5.6.1.4 Inclusion of Unplanned Heart Failure Visits

This analysis will be identical to the ITT analysis with the exception that in addition to recurrent HF hospitalizations, Unplanned HF Visit events will be included as defined in the protocol.

An "unplanned heart failure visit" is considered to include unscheduled office visits or Emergency Department (ED) visits that meet the definition for HF hospitalization, with the exception of the 24-hour requirement.

5.6.1.5 Andersen-Gill Counting Process Analysis

The Andersen-Gill (A-G) counting process method will also be used to model the recurrent hospitalization data and estimate the RRR as a sensitivity analysis.

A robust estimate of the variance will be applied in the calculation of confidence intervals and tests of hypotheses.

5.6.1.6 Tipping-point Analysis

Tipping-point analysis on the primary effectiveness endpoint will be conducted as a sensitivity analysis. Subjects in the Device or Control group who withdraw from the study before completing 24-month follow-up visit will be analyzed one at a time as having experienced a HF hospitalization on the date of withdrawal.

5.6.1.7 Andersen-Gill Model with a Time Varying Covariate

The A-G model assumes proportional hazards. However, it is possible that the hazard of subsequent HF hospitalization increases if the subject has experienced a prior HF hospitalization. Accordingly, the A-G model will be re-fit with a time varying indicator to account for a prior HF related hospitalization post randomization.

5.6.1.8 Andersen-Gill Counting Process Analysis for Composite of Recurrent Heart-Failure Hospitalization or Death

The Andersen-Gill (A-G) counting process method will also be used to model the composite of recurrent HF hospitalization or death. The RRR and 95% confidence interval calculated from the A-G model will be provided as a sensitivity analysis. A robust estimate of the variance will be applied in the calculation of confidence intervals and tests of hypotheses.

5.6.1.9 Multi-State Model

A multi-state model[2] will be fit to the effectiveness data as a sensitivity analysis. The model will include four states corresponding to:

- State 1: "Treatment" visit
- State 2: 1st HF hospitalization
- State 3: Additional HF hospitalization
- State 4: Death

The transitions between the states will be estimated using a stratified proportional hazards regression model with a treatment group indicator (Device, Control) and with strata corresponding to the six transitions {1→2, 1→4, 2→3, 2→4, 3→3, 3→4}. The homogeneity of the Device effect on HF hospitalizations and death will be evaluated via a test of the treatment by strata interaction. A likelihood ratio test based on the difference between the full model (interaction) and reduced model (no interaction) likelihood ratio will be used. The estimate of the Device effect will be based on the hazard ratio of the Device vs. Control.

5.6.1.10 Subgroup Analyses for Primary Effectiveness Endpoint

5.6.1.10.1 Sex (male vs. female)

Subgroup analyses of sex will be performed. Major baseline and clinical characteristics will be summarized by sex, and by treatment group.

The JFM model will be re-fit with the effects of treatment group, sex and the interaction of treatment group by sex. The test for the interaction effect on the primary effectiveness endpoint will be performed at the 15% level of significance.

5.6.1.10.2 Etiology of Cardiomyopathy (ischemic vs. non-ischemic)

Subgroup analyses of cardiomyopathy etiology will be performed. Major baseline and clinical characteristics will be summarized by etiology of cardiomyopathy, and by treatment group.

The JFM model will be re-fit with the effects of treatment group, cardiomyopathy etiology and the interaction of treatment group by etiology. The test for the interaction effect on the primary effectiveness endpoint will be performed at the 15% level of significance.

5.6.1.10.3 Left Ventricular Ejection Fraction (LVEF>40% vs. LVEF≤40%)

Subgroup analyses of Central Echocardiography Core Laboratory assessed baseline LVEF will be performed. Major baseline and clinical characteristics will be summarized by baseline LVEF and by treatment group.

The JFM model will be re-fit with the effects of treatment group, EF and the interaction of treatment group by EF. The test for the interaction effect on the primary effectiveness endpoint will be performed at the 15% level of significance.

5.6.1.10.4 Extreme Surgical Risk Status

Subgroup analyses of extreme surgical risk status as assessed by the Central Eligibility Committee will be performed. Major baseline and clinical characteristics will be summarized by extreme surgical risk status, and by treatment group.

The JFM model will be re-fit with the effects of treatment group, extreme surgical risk status and the interaction of treatment group by extreme surgical risk status. The test for the interaction effect on the primary effectiveness endpoint will be performed at the 15% level of significance.

5.6.1.11 Analysis of Poolability of Study Sites

Analysis of the site effect will be performed. The A-G model will include the effects of treatment received, site, and the interaction of treatment by site. The interaction test will be performed at the 15% level of significance. The study will involve up to 100 sites. It is likely that there will be wide variation in the number of subjects enrolled across sites (due to differences in site start-up times and differences in subject volumes enrolled at sites). If model convergence is not achieved, sites will be categorized into the categories described below. The model will then include the effects of treatment, site category and the interaction

of treatment by site category. In addition, a descriptive summary of recurrent HF hospitalizations will be provided by site.

- Large: Sites with more than 10 subjects
- Medium Large: Sites that enroll between 6 and 10 subjects
- Medium: Sites that enroll between 3 and 5 subjects
- Small: Sites that enroll fewer than 3 subjects.

5.6.2 Secondary Effectiveness Endpoints

The secondary effectiveness endpoints will be tested at a two-sided significance level of 5%, unless otherwise specified. This primary analysis population of the secondary effectiveness endpoints will be the ITT population. Additional analyses of the secondary effectiveness endpoints will be performed on the PP and AT populations.

5.6.2.1 MR Severity at 12 Months

Subjects in the Device group are expected to experience greater reduction in MR severity than subjects in the Control group.

Hypothesis:

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $P_{D,MR\leq 2+} - P_{C,\ MR\leq 2+} = 0$ vs.

$\mathbf{H_1}$: $P_{D,\ MR\leq 2+} - P_{C,\ MR\leq 2+} \neq 0$

where $P_{D,\ MR\leq 2+}$ and $P_{C,\ MR\leq 2+}$ represent the proportion of subjects with MR severity ≤ 2+ at 12 months in the Device and Control groups, respectively.

Fisher's exact test at the 5% level of significance will be used to compare the proportion of subjects in the Device and Control groups with MR severity ≤ 2+ at 12 months.
As a sensitivity analysis, for subjects who experience an adjudicated heart failure death prior to completing 12 months follow-up, these subjects will be classified with MR of severity 4+ in the analysis. This sensitivity analysis is exploratory and Type I error control does not apply to these hypotheses.

5.6.2.2 Change in Six-Minute Walk Test Distance (6MWD) at 12 Months over Baseline

Improvement in the 6MWD at 12 months from baseline is an important secondary effectiveness endpoint in the comparison of the Device group to the Control group.

Hypothesis:

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $\mu_{T,\Delta 6MWD} - \mu_{C,\Delta 6MWD} = 0$ vs.

$\mathbf{H_1}$: $\mu_{T,\Delta 6MWD} - \mu_{C,\Delta 6MWD} \neq 0$

where $\mu_{T,\Delta 6MWD}$ and $\mu_{C,\Delta 6MWD}$ represent the mean change in distance walked between 12 months and baseline in the Device and Control groups respectively.

The analysis will be performed in subjects with paired 6MWD data at baseline and 12 months. Subjects who experience adjudicated heart failure death prior to completing 12 months follow-up or are unable to exercise due to cardiac reasons will also be included in the analysis. These subjects will be assigned 6MWD of 0 meters at 12 months in the analysis.

Analysis of covariance (ANCOVA) will be used to compare the mean changes in 6MWD between 12 months and baseline in the two groups, while adjusting for baseline 6MWD value. Least squares means (LS-Means) and 95% confidence interval for the change in 6MWD between 12 months and baseline will be calculated for the Device and Control groups, respectively.

A two-sample t-test and a Wilcoxon Rank Sum test will be conducted as supplementary analyses.

In addition to a significant p-value for this analysis, the point estimate of the difference in mean improvement between the two groups (Device-Control) should be at least 24 meters to meet this endpoint.

5.6.2.3 Change in Quality of Life (Kansas City Cardiomyopathy Questionnaire, KCCQ) at 12 Months over Baseline

Improvement in quality of life as measured by the KCCQ Test at 12 months from baseline will be a secondary effectiveness endpoint in the comparison of the Device group to the Control group.

Hypothesis:

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $\mu_{D,\Delta KCCQ} - \mu_{C,\Delta KCCQ} = 0$ vs.

$\mathbf{H_1}$: $\mu_{D,\Delta KCCQ} - \mu_{C,\Delta KCCQ} \neq 0$

where $\mu_{D,\Delta KCCQ}$ and $\mu_{C,\Delta KCCQ}$ represent the mean change in quality of life score between 12 months and baseline in the Device and Control groups respectively.

The analysis will be performed in subjects with paired KCCQ score data at baseline and 12 months. Subjects who experience an adjudicated heart failure related cardiovascular death prior to completing 12 months follow-up will also be included in the analysis. These subjects will be assigned the lowest KCCQ score at 12 months observed for any subject in the analysis. The lowest possible KCCQ score is 0 point.

ANCOVA will be used to compare the mean changes in KCCQ score between 12 months and baseline in the two groups, while adjusting for baseline KCCQ score. LS-Means and 95% confidence interval for the change in KCCQ score between 12 months and baseline will be calculated for the Device and Control groups, respectively.

A two-sample t-test and a Wilcoxon Rank Sum test will be conducted as supplementary analyses.

In addition to a significant p-value for this analysis, the point estimate of the difference in mean improvement between the two groups (Device-Control) should be at least 5 units to meet this endpoint.

5.6.2.4 Change in LVEDV at 12 Months over Baseline

Reduction in MR severity with the MitraClip device is expected to be associated with reduction in LVEDV.

Hypothesis:

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $\mu_{D,\Delta LVEDV} - \mu_{C,\Delta LVEDV} = 0$ vs.

$\mathbf{H_1}$: $\mu_{D,\Delta LVEDV} - \mu_{C,\Delta LVEDV} \neq 0$

where $\mu_{D,\Delta LVEDV}$ and $\mu_{C,\Delta LVEDV}$ represent the mean change in left ventricular end diastolic volumes at 12 months over baseline in the Device and Control groups, respectively.

All subjects with paired LVEDV data at baseline and 12 months will be included in the analysis. Subjects who experience an adjudicated heart failure death prior to completing 12 months follow-up will also be included in the analysis. These subjects will be assigned the worst LVEDV change at 12 months from baseline observed for any subject in the analysis.

ANCOVA will be used to compare the mean changes in LVEDV between 12 months and baseline in the two groups, while adjusting for baseline LVEDV score. LS-Means and 95% confidence interval for the change in LVEDV between 12 months and baseline will be calculated for the Device and Control groups, respectively.

A two-sample t-test and a Wilcoxon Rank Sum test will be conducted as supplementary analyses.

5.6.2.5 NYHA Functional Class I or II at 12 months

Reduction in MR severity with the MitraClip device is expected to be associated with higher proportion of NYHA Functional Class I or II at 12 months in the Device group.

The null and alternative hypotheses may be stated as:

$\mathbf{H_0}$: $P_{D,NYHA\ I/II} - P_{C,NYHA\ I/II} = 0$ vs.

$\mathbf{H_1}$: $P_{D,NYHA\ I/II} - P_{C,NYHA\ I/II} \neq 0$

where $P_{D,NYHA\ I/II}$ and $P_{C,NYHA\ I/II}$ represent the proportion of subjects with NYHA Class I/II at 12 months in the Device and Control groups, respectively.

Fisher's exact test at the 5% level of significance will be used to compare the proportion of subjects in the Device and Control groups with NYHA Class I/II at 12 months.

All subjects with NYHA data at 12 months will be included in the analysis. Subjects who experience an adjudication heart failure death prior to completing 12 months follow-up will also be included in the analysis. These subjects will be assigned a 12-month NYHA class of IV.

5.6.2.6 Finkelstein-Schoenfeld Analysis of All-Cause Death and Recurrent HF Hospitalization

The hierarchy of all-cause death and recurrent HF hospitalization will be analyzed using the Finkelstein-Schoenfeld[3] method. Hospitalizations that are adjudicated by the CEC as related to HF using the pre-specified protocol definition will be included in the analysis.

The null and alternative hypotheses are stated as:

$\mathbf{H_0}$: $\lambda_{D,Death} = \lambda_{C,Death}$ **AND** $\lambda_{D,Hosp} = \lambda_{C,Hosp}$

$\mathbf{H_1}$: $\lambda_{D,Death} \neq \lambda_{C,Death}$ **OR** $\lambda_{D,Hosp} \neq \lambda_{C,Hosp}$

Where $\lambda_{D,Death}$ and $\lambda_{C,Death}$ represent the rate of all-cause mortality in the Device and Control groups respectively. $\Lambda_{D,Hosp}$ and $\lambda_{C,Hosp}$ represent the rate of recurrent HF hospitalizations in the Device and Control groups respectively.

This endpoint will be analyzed when the last subject completes 12 months of follow-up. All available follow-up through 24 months at the time of data cut-off will be included in the analysis. These composite hypotheses will be analyzed using the method of Finklestein and Schoenfeld with death higher in the hierarchy than recurrent HF hospitalizations.

Specifically, for each pair of subjects (say subjects *i* and *j*), define a score u_{ij} in the following manner:

- If subject *i* is known to have experienced death at a later date than subject *j*, then $u_{ij} = 1$ (if subject *j* is known to have experienced death at a later date, then $u_{ij} = -1$). This determination would happen if death dates are available for both subjects, or if one subject was censored at a later time than the death time for the other.

- If it is not known which subject has gone longer without death, then compare the number of recurrent HF hospitalizations for the two subjects within the shorter of the two follow-up durations. If subject *i* has fewer hospitalizations than subject *j*, then $u_{ij} = 1$; (if subject *j* is known to have fewer hospitalizations, then $u_{ij} = -1$). If subject *i* has same number of hospitalizations as subject *j*, then compare the time to the first hospitalization. If subject *i* is known to have the first hospitalization at a later date than subject *j*'s first hospitalization, then $u_{ij} = 1$ (if subject *j* is known to have the first hospitalization at a later date than subject *i*'s first hospitalization, then $u_{ij} = -1$)

- For ties, $u_{ij} = u_{ji} = 0$. And in all cases, $u_{ij} = -u_{ji}$.

Note that the score looks first for a difference in freedom from death. If there is no difference in this measure, then the score looks for difference in the rate and timing of recurrent hospitalizations for HF. The final test statistic is based on the sum of the scores for subjects in the Device group. If we let $D_i = 1$ for subjects in Device group and let $D_i = 0$ for subjects in the Control group, we define the statistic using the score described above:

$$T = \sum_{i=1}^{n} U_i D_i$$

where

$$U_i = \sum_{i \neq j} u_{ij}$$

The mean of the test statistic is zero under the null hypothesis of no difference between Device and Control. Finkelstein and Schoenfeld derive the variance for this statistic:

$$V = \frac{n_D(n - n_D)}{n(n-1)} \sum_{i=1}^{n} U_i^2$$

where, n_D is the number of subjects in the Device group. The null hypothesis will be rejected at the two-sided significance level of 5% favoring the Device group if $\frac{T}{V^{1/2}}$ is greater than the upper 97.5th percentile of the standard normal distribution**.**

Furthermore, the unmatched win-ratio[8] approach will be used to evaluate the composite endpoint of death and recurrent HF hospitalizations through 24 months at the time of data cut-off.

Specifically, each pair of subjects (say subjects *i* and *j*) can be classified into one of categories (a), (b), (c), (d), or (e):

(a) Death in the Device group first.

(b) Death in the Control group first.

(c) More HF hospitalizations in the Device group (in the case of tie, the first HF hospitalization in the Device group occurs first).

(d) More HF hospitalization in the Control group (in the case of tie, the first HF hospitalization in the Control group occurs first).

(e) None of the above.

The composite endpoint results are summarized by N_a, N_b, N_c, N_d, N_e, the number of pairs in categories (a), (b), (c), (d), and (e), respectively. Based on death and HF hospitalization, respectively, $N_W = N_b + N_d$ is the number of "winners" for the Device group while $N_L = N_a + N_c$ is the number of "losers" for the Device group. The "win ratio" is defined as the number of "winners" divided by the number of "losers", i.e., $R_w = N_w/N_L$.

The estimation of the win ratio and 95% confidence interval will be provided to estimate the treatment effect of the composite endpoint of all-cause death and HF hospitalizations through 24 months.

5.6.2.7 Recurrent Hospitalization - all-cause

Treatment with the MitraClip device may reduce the risk of recurrent all-cause hospitalization.

Hypothesis:

The null and alternative hypotheses are stated as:

H_0: RRR = 0 vs.

H_1: RRR ≠ 0

where RRR is the relative risk reduction in the rate of recurrent hospitalization due to treatment with the MitraClip device. All hospitalizations will be included as events in the analysis.

The Joint Frailty Model method will be used to model the recurrent hospitalization data and estimate the RRR.

RRR = (1- Hazard Ratio of Device vs. Control) = $(1-e^{\beta})$ where β is the regression coefficient of the recurrent hospitalization in the JFM model.

This endpoint will be analyzed when the last subject completes 12 months of follow-up. All available follow-up through 24 months at the time of data cut-off will be included in the analysis. Subjects who do not experience any hospitalizations will be censored on their data cut-off date. Subjects who die will be censored on the date of death. Subjects who withdraw from the trial without a hospitalization will be censored on the date of withdrawal.

The null hypothesis is rejected if the regression coefficient for the Device group is significant at the two-sided significance level of 5%. The two-sided 95% confidence interval is calculated as RRR ± $Z_{\alpha/2}$ × SE where RRR and SE are estimated from the Joint Frailty Model and $Z_{\alpha/2} = 100(1-\alpha/2)$th percentile of the standard normal distribution. Beneficial treatment effect is established if the lower two-sided 95% confidence bound of the relative risk reduction is greater than zero.

5.7 Analysis of Additional Descriptive Endpoints

All additional endpoints in this section will be reported with descriptive statistics. Continuous endpoints will be summarized using the following: mean, standard deviation, median, the first and third quartiles, range, and the 95% confidence interval for the population mean using the normal approximation. Categorical endpoints will be summarized as proportions in the form of fractions and percent. The 95% confidence interval for a population proportion will be constructed using the Clopper-Pearson exact method[6]. The 95% confidence interval for the difference of two proportions will be constructed using the Newcombe [7]score method. Other analyses are specified under each category of additional descriptive endpoints.

5.7.1 Device or Procedure-Related Adverse Events

Device or procedure-related adverse events are defined as adverse events that are adjudicated by the Clinical Events Committee as possibly, probably or definitely device and/or procedure-related, regardless of the temporal relationship to the MitraClip procedure. Device or procedure-related adverse events will be broken down into those that occur within 30 days of the procedure and those that occur after 30 days of the procedure. Examples of device-related adverse events are:

- Myocardial perforation
- Single Leaflet Device Attachment
- Embolization of the MitraClip device or MitraClip System components
- Iatrogenic atrial septal defect
- Mitral valve stenosis
- Need for mitral valve replacement instead of repair due at least in part to the MitraClip procedure or the presence of the MitraClip device

5.7.2 Device and Procedure-Related Endpoints

The following device and procedure-related acute endpoints will be reported for the Device group:

- Implant Rate: defined as the rate of successful delivery and deployment of MitraClip device with echocardiographic evidence of leaflet approximation and retrieval of the delivery catheter
    - Any additional MitraClip device implantation after the index procedure will be summarized and reported as an additional MitraClip device intervention and will be further reported as a device or procedure failure, if appropriate.
- Device Procedure Time: defined as the time elapsed from the start of the transseptal procedure to the time the Steerable Guide Catheter is removed
- Total Procedure Time: defined as the time elapsed from the first of any of the following: intravascular catheter placement, anesthesia or sedation, or transesophageal echocardiogram (TEE), to the removal of the last catheter and TEE.
- Device Time: defined as the time the Steerable Guide Catheter is placed in the intra-atrial septum until the time the MitraClip Delivery System (CDS) is retracted into the Steerable Guide Catheter.
- Fluoroscopy duration: defined as the duration of exposure to fluoroscopy during the MitraClip procedure

5.7.3 Echocardiographic Endpoints

The following echocardiographic endpoints will be reported for the Device and Control groups at baseline, discharge (or 30 days if discharge echocardiogram is not available), 6 months,12 months, 24 months and then yearly through 5 years. For continuous variables, change from baseline to each follow-up will also be reported:

- MR Severity GradeEffective Regurgitant Orifice Area
- Regurgitant Volume

- Regurgitant Fraction
- Left Ventricle End Diastolic Volume (LVEDV)
- Left Ventricular End Systolic Volume (LVESV)
- Left Ventricular End Diastolic Dimension (LVEDD)
- Left Ventricular End Systolic Dimension (LVESD)
- Left Ventricular Ejection Fraction (LVEF)
- Right Ventricular Systolic Pressure (RVSP)
- Mitral Valve Area
- Mean Mitral Valve Gradient
- Systolic Anterior Motion of the mitral valve (present or absent)
- Cardiac Output
- Forward Stroke Volume

5.7.4 Clinical Endpoints

The following clinical endpoints will be reported for the Device and Control groups. For endpoints where a baseline measurement and post-treatment measurements are collected, the change from baseline will be summarized by time-point for each group and compared between Control and Device groups.

- Kaplan-Meier freedom from the components of the primary safety composite at 12 months, 24 months and yearly through 5 years (Device group only). Deaths and withdrawals will be censored in the analysis
- Kaplan-Meier freedom from the primary safety composite at 24 months and yearly through 5 years (Device group only)
- Kaplan-Meier freedom from all-cause mortality at 12 months, 24 months and then yearly through 5 years
- Kaplan-Meier freedom at 12 months, 24 months and then yearly through 5 years from:
    1) cardiovascular mortality
    2) the first HF related hospitalization

3) the first cardiovascular hospitalization

4) the first HF related hospitalization or all-cause mortality

- Proportion of MR Severity Grade < 2+ at 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years.
- NYHA Functional Class at baseline, 30 days, 6 months,12 months, 24 months and then yearly through 5 years
- Proportion of >1 class improvement of NYHA Functional Class at 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years from baseline.
- Mean change in NYHA Functional Class at 30 days, 6 months, 12 months, 24 months and then yearly through 5 years from baseline.
- 6MWD at baseline, 30 days,6 months,12 months and 24 months (and change from baseline to follow-up)
- KCCQ QoL scores at baseline, 30 days,6 months, 12 months and 24 months (and change from baseline to follow-up)
- SF-36 QoL scores at baseline, 30 days, 6 months, 12 months and 24 months (and change from baseline to follow-up)
- Mitral valve surgery (including type of surgery), new use of CRT, new use of single or dual chamber pacemaker, permanent LVAD implant, heart transplant, additional MitraClip device intervention in Device group or *de novo* MitraClip device intervention in Control group, including reason for intervention through 5 years
- Responder analysis for 6MWD, where responder is defined as alive and experiencing an improvement of 24 meters and 50 meters (difference in proportion of responders between Device and Control groups) at 12 months and 24 months
- Responder analysis for LVEDV Index, where responder is defined as alive and experiencing an improvement of 12 ml/m^2 (difference in proportion of responders between Device and Control groups) at 12 months, 24 months and then yearly through 5 years
- Responder analysis for QoL (KCCQ), where responder is defined as alive and experiencing an improvement of 5 points (difference in proportion of responders between Device and Control groups) at 12 months and 24 months
- Each subscale for QoL (KCCQ) (difference in means between Device and Control groups) at 12 months and 24 months
- Length of index hospitalization for MitraClip procedure (Device group)

- Number of hospitalizations and reason for hospitalizations (i.e., HF, cardiovascular, non-cardiovascular) at 12 months and 24 months in each of the Device and Control groups

- Number of days alive and out of hospital from the time of randomization (difference in medians between Device and Control groups) to 12 months, 24 months and then yearly through 5 years

- Number of days hospitalized from the "Treatment" visit (difference in medians between Device and Control groups) at 12 months, 24 months and then yearly through 5 years

- Proportion of alive time in hospital will be summarized and compared between Device and Control groups at 12 months, 24 months and then yearly through 5 years

- Proportion of subjects living in the baseline location at 12 months, 24 months and then yearly through 5 years

- Mitral valve replacement rates will be summarized and compared between Device and Control groups at 12 months, 24 months and then yearly through 5 years

- New onset of permanent atrial fibrillation at 12 months, 24 months and then yearly through 5 years

- Mitral stenosis at 12 months, 24 months and then yearly through 5 years

- Clinically significant atrial septal defect (ASD) that requires intervention at 12 months,24 months and then yearly through 5 years

- Device-related complications in Device group subjects and Control group subjects who undergo MitraClip procedure through 5 years

- BNP (Brain Natriuretic Peptide) or NT-pro BNP (N-terminal prohormone of Brain Natriuretic Peptide) levels at baseline, 30 days, and 12 months

- Modified Rankin Scale Score at baseline, 30 days, 6 months, 12 months

- Major bleeding at 30 days

- Prolonged ventilation at 30 days

- Compliance with Guideline Directed Medical Therapy (GDMT) at baseline, 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years

- Average dosages of Guideline Directed Medical Therapy (GDMT) at baseline, 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years

- The number of (1) any changes in GDMT and GDMT dosage from baseline at 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years and (2) any changes in GDMT from baseline that result in a larger than 100% increase or 50% decrease in dose at 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years

- The reasons for any changes in GDMT from baseline at 30 days, 6 months, 12 months, 24 months, and then yearly through 5 years

5.7.5 CPX Sub-Study Endpoint

A sub-study endpoint will be added utilizing peak VO_2 as a parameter for cardiopulmonary exercise testing (CPX). CPX testing will be conducted at baseline and 12 months on a total of at least 50 and up to 100 subjects at qualified sites. Mean changes in peak VO_2 (ml/kg/min) will be summarized at 12 months from baseline for the subset of patients who complete a CPX test at baseline and 12 months. A comparison of change from baseline between Device and Control groups will be presented. A p-value for the peak VO_2 endpoint will be generated for descriptive purposes.

5.7.6 Centers for Medicare and Medicaid Services (CMS) Endpoints

Additional clinical outcomes will be performed for the Centers for Medicare and Medicaid Services (CMS) requirements for the COAPT Trial. The analysis will be performed at the time of the COAPT trial's primary endpoints analyses and these data will be reported separately to the CMS Coverage and Analysis Group at the same time the results are submitted to the Food and Drug Administration (FDA). Refer to Appendix A for the details.

5.8 Power and Sample Size

A total sample size of 610 subjects is planned. The power for the primary safety endpoint when 305 Device group subjects complete 12 months of follow-up is > 95%. The power for the primary effectiveness endpoint with follow-up through 24 months in 610 subjects when the last subject completes 12 months of follow-up is ~ 80%. The overall power for the study for the primary endpoints with the planned randomization of 610 subjects is ~80%.

5.8.1 Primary Safety Endpoint

Table 1 lists the event rates assumed for the components of the composite primary safety endpoint in the Device group

Table 1: Assumptions for Primary Safety Event Rates at 12 Months

Event	Device Group
SLDA	4%
Embolization	<0.1%
Endocarditis	<0.1%
Mitral Stenosis Confirmed by ECL That Requires Surgery	<0.1%
LVAD Implant, Heart Transplant	1%
Other Device-related Complications That Require Non-elective Cardiovascular Surgery	<0.1%
Total	~6%

Two thousand (2000) simulations were performed to calculate sample size and power for the primary safety endpoint. Assuming 22% mortality and 7.5% attrition at 12 months, a total of 305 subjects in the Device group will provide > 95% power to reject the null hypothesis at the one-sided significance level of 5%.

5.8.2 Primary Effectiveness Endpoint

At least ten thousand (10,000) simulations with converged joint frailty model were performed to calculate sample size and power for the primary effectiveness endpoint. With a fixed sample size design, 610 subjects provide approximate 80% power at the one-sided significance level of 5% for the primary effectiveness endpoint.

The following assumptions were made for the calculation of sample size:

- The enrollment rate is based on the actual monthly enrollment numbers from December 2012 to February 2016, then 20 subjects per month, thereafter
- Attrition due to withdrawals at 12 months is 7.5% in both groups
- In the EVEREST High Risk Cohort (N = 211), the lower bound of a 95% confidence interval on the HF hospitalization rate in the period pre-enrollment was ~0.6 per 12 months. The HF hospitalization rate in the Control group is therefore assumed to be 0.6 per 12 months. The HF hospitalization rate in the Device group is assumed to be 0.42 per 12 months (30% Relative Risk Reduction).
- Subjects are followed for a minimum of 12 months. All available follow-up through 24 months at the time of data cut-off will be included in the analysis.
- The mortality rate in the Control group is 27%, and in the Device group is 22%.
- To account for informative censoring of death event, a shared random effect (frailty) model for recurrent events and terminal event is used to simulate recurrent events and

death event data. The correlation between these two processes is introduced by the shared frailty distribution, Gamma distribution with mean of 1 and variance of 1. The frailty is assumed to have the same impact on the hazards for recurrent and terminal event (e.g. frailty exponent is 1). The frailty variance and frailty exponent values are estimated and determined based on the REALISM HR FMR patients data and blinded pooled COAPT roll-in and randomized data as of June, 2016. The key underlying assumptions of the JFM are

a) The recurrent, terminating, and censoring processes all have continuous distribution so that recurrent events and death cannot happen at the exact same time.

b) Death event precludes the observation of new recurrent HF hospitalization events, but independent censoring (e.g. withdrawal or lost-to- follow-up) does not interrupt the occurrence of recurrent HF hospitalization events.

c) The intensity of the current event process at time *t* is a function of subject's process history, covariates, frailty, and being alive just before time *t*. The intensity of the death process at time *t* is a function of covariates, frailty and being alive just before time *t*. Both recurrent process and death process are correlated through the same frailty distribution.

d) Censoring (administrative censoring) is noninformative, which does not depend on unobserved frailty.

Detailed information about the joint frailty model assumptions, model specification and analysis method and assumed frailty parameters can be found in a separate appendix to the SAP.

- The Joint Frailty Model method was used to estimate the logarithm of 1-RRR for recurrent HF hospitalization event.

5.8.3 Secondary Safety Endpoints

5.8.3.1 Composite 30-Day Secondary Safety Endpoint

Table 2 lists the event rates assumed for the components of the composite secondary safety endpoint in the Device group.

Table 2: Assumptions for Secondary Safety Event Rates at 30 Days

Event	Device Group
Death	5%
Stroke	2%
MI	2%
Non-elective CV Surgery for Device Related Complications	3%
Total	12%

Two thousand (2000) simulations were performed. A total of 305 subjects in the Device group (without attrition) provides > 95% power to reject the null hypothesis at the one-sided significance level of 5%.

5.8.3.2 All-Cause Mortality at 12 Months

Two thousand (2000) simulations were performed. Assuming 27% mortality in the Control group, 22% mortality in the Device group, and 7.5% attrition in both groups at 12 months, a total of 610 subjects will provide > 90% power to reject the null hypothesis at the one-sided significance level of 5%.

5.8.4 Secondary Effectiveness Endpoints

Unless otherwise specified, the power calculations of the following secondary effectiveness endpoints are based on the two-sided significance level of 5%.

5.8.4.1 Change in KCCQ QoL score at 12 Months

A simulation was performed. Based on the following assumptions, the power is > 90% to reject the null hypothesis at the 5% significance level.

- Subjects in the Device group have 8 points more improvement in the mean change of KCCQ score at 12 months from baseline than subjects in the Control group, with standard deviation (SD) of 25 points for the mean change in KCCQ score in each group.

- Assume all-cause mortality rate at 12 months is 27% in the Control group, and 22% in the Device group. Approximate 70% of all-cause deaths in each group are due to cardiovascular reasons. Furthermore, about 70% of cardiovascular deaths are HF related. Thus, it is expected to have approximate 3% difference in HF related cardiovascular deaths at 12 month between two treatment groups (i.e. 10% in the Device group and 13% in the Control group). Subjects who experienced heart failure death prior to 12 months follow-up are assigned with the lowest KCCQ score at 12 months.

- 10% attrition rate in both groups due to withdrawn or missed visit at 12 months.
- Two-sample *t*-test (it is anticipated that an ANCOVA analysis adjusting for baseline values provides as much as or more power than a two-sample *t*-test [9,10]).

5.8.4.2 MR severity at 12 Months

Sample size calculation for the MR severity at 12 months was performed using NCSS PASS 11 (Hintze, J., 2011, NCSS, LLC. Kaysville, Utah).

The power calculation for the MR severity at 12 months is based on the following assumptions:

- At least 20% more subjects in the Device group had MR ≤ 2+ at 12 months than those subjects in the Control group.
- Assume all-cause mortality rate at 12 months is 27% in the Control group, and 22% in the Device group.
- 20% attrition rate in both groups due to withdrawn, missed visit, or unevaluable echocardiography at 12 months.
- Fisher's exact test at the 5% level of significance

With the effective sample size of total 300 subjects at 12 months, there is >90% power to demonstrate that the Device group has achieved more MR ≤2+ at 12 months than the Control group.

5.8.4.3 Hierarchy of death and recurrent HF hospitalization (analyzed when the last subject completes 12 months of follow-up)

A simulation was performed. Based on the same assumptions as stated from the primary effectiveness endpoint, and Finkelstein-Schoenfeld Analysis method for the hierarchy of death and recurrent HF hospitalization, a total sample size of 610 subjects provides approximate 84% power to detect 10% absolute difference in all-cause mortality (i.e. 29% for the Control group and 19% for the Device group) and 30% relative risk reduction in the HF hospitalization (i.e. The HF hospitalization rate per 12 months is 0.6 in the Control group and 0.42 in the Device group).

5.8.4.4 Change in 6MWD at 12 Months

A simulation was performed. Based on the following assumptions, the power is approximate 90% to reject the null hypothesis at the 5% significance level.

- Subjects in the Device group have 30 meters more improvement in the mean change in 6MWD at 12 months from baseline than subjects in the Control group, with standard deviation (SD) of 100 meters for the mean change in 6MWD in each group.

- Assume all-cause mortality rate at 12 months is 27% in the Control group, and 22% in the Device group. Approximate 70% of all-cause deaths in each group are due to cardiovascular reasons. Furthermore, about 70% of cardiovascular deaths are HF related. Thus, it is expected to have approximate 3% difference in HF related cardiovascular deaths at 12 month between two treatment groups (i.e. 10% in the Device group and 13% in the Control group). Subjects who experienced heart failure death prior to 12 months follow-up are assigned with 0 meter of 6MWD at 12 months.

- 10% attrition rate in both groups due to withdrawn or missed visit at 12 months.

- Two-sample *t*-test (it is anticipated that an ANCOVA analysis adjusting for baseline values provides as much as or more power than a two-sample *t*-test).

5.8.4.5 NYHA Functional Class of I/II at 12 months

Sample size calculation for the NYHA Functional Class I/II at 12 months was performed using NCSS PASS 11 (Hintze, J., 2011, NCSS, LLC. Kaysville, Utah).

The power calculation for the NYHA Functional Class I/II rate at 12 months is based on the following assumptions:

- Assume 60% of subjects in the Control group had NYHA Functional Class I/II at 12 months, and 75% of subjects in the Device group had NYHA Functional Class I/II at 12 months. Thus, there is 15% more NYHA Functional Class of I/II at 12 months in the Device group than those in the Control group.

- Assume all-cause mortality rate at 12 months is 27% in the Control group, and 22% in the Device group. Approximate 70% of all-cause deaths in each group are due to cardiovascular reasons. Furthermore, about 70% of cardiovascular deaths are HF related. Thus, it is expected to have approximate 3% difference in HF related cardiovascular deaths at 12 month between two treatment groups (i.e. 10% in the Device group and 13% in the Control group). Subjects who experienced heart failure death prior to 12 months follow-up are assigned with NYHA Functional Class IV at 12 months.

- 10% attrition rate in both groups due to withdrawn or missed visit at 12 months.

- Fisher's exact test

With the effective sample size of approximate 220 subjects in each arm at 12 months, there is approximate 90% power to demonstrate that there are more subjects in the Device group who had NYHA Functional Class I/II at 12 months than those in the Control group.

5.8.4.6 All-cause Mortality through 24 months (superiority)

A simulation was performed. Assuming the all-cause mortality rate at 12 months is 29% in the Control group, and 19% in the Device group, and 7.5% yearly attrition rate due to withdrawals in both groups, and the constant hazard ratio up to 24 months between two treatment groups, a total of 610 subjects will provide approximate 80% power to detect 10% all-cause mortality difference at 12 months.

5.8.4.7 Change in Left Ventricular End Diastolic Volume at 12 Months

A simulation was performed. Based on the following assumptions, the power is approximate 80% to reject the null hypothesis at the 5% significance level.

- Assume that subjects in the Device group have 12 mL more reduction in the mean change of LVEDV than subjects in the Control group, with standard deviation (SD) of 45 mL for the change in LVEDV in each group.
- Assume all-cause mortality rate at 12 months is 27% in the Control group, and 22% in the Device group. Approximate 70% of all-cause deaths in each group are due to cardiovascular reasons. Furthermore, about 70% of cardiovascular deaths are HF related. Thus, it is expected to have approximate 3% difference in HF related cardiovascular deaths at 12 month between two treatment groups (i.e. 10% in the Device group and 13% in the Control group). Subjects who experienced heart failure death prior to 12 months follow-up are assigned with the worst observed change in LVEDV at 12 months from baseline.
- 20% attrition rate in both groups due to withdrawn, missed visit, or unevaluable echocardiography at 12 months.
- Two-sample *t*-test (it is anticipated that an ANCOVA analysis adjusting for baseline values provides as much as or more power than a two-sample *t*-test).

5.8.4.8 All-cause Recurrent Hospitalizations

A simulation was performed. Based on the same assumptions as stated from the primary effectiveness endpoint of recurrent HF hospitalization, a total sample size of 610 subjects provides approximate 80% power to detect 26% relative risk reduction in the HF hospitalization (i.e. the all-cause hospitalization rate per patient per 12 months is 1.9 in the Control group and 1.4 in the Device group). In the simulation, the all-cause mortality rates at 12 months are assumed to be 27% for the Control group and 22% for the Device group.

5.9 Multiplicity Adjustments

Both the primary safety and primary effectiveness endpoints must be met for trial success.

The secondary endpoints will be evaluated for labeling claims if both primary endpoints are met. The secondary endpoints will be tested sequentially in the following order:

- MR Severity of ≤ 2+ at 12 months
- All-Cause Mortality at 12 months (Non-inferiority test)
- Hierarchy of death and recurrent HF hospitalization (analyzed when the last subject completes 12 months of follow-up)
- Change in KCCQ score at 12 months from baseline
- Change in 6MWD at 12 months from baseline
- Recurrent hospitalizations - all-cause (analyzed when the last subject completes 12 months of follow-up)
- NYHA Functional Class of I/II at 12 months
- Change in Left Ventricular End Diastolic Volume at 12 months
- All-Cause Mortality (Superiority test, if HR < 1) (analyzed when the last subject completes 12 months of follow-up and use all available data up to 24 months)
- Composite 30-Day Secondary Safety Endpoint

The multiplicity adjustments above are only performed for the primary analyses (ITT) of the primary and secondary endpoints.

5.10 Trial Success

Both the primary safety and primary effectiveness endpoints must be met for trial success. Additional labeling claims may be made based on the secondary endpoints.

6 References

1. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 2007;8(4):708-21

2. H. Putter, M. Fiocco, and R.B. Geskus Tutorial in biostatistics: Competing risks and multi-state models, *Statistics in Medicine* 2007;26: 2389-2430

3. Dianne M. Finkelstein, David A. Schoenfeld, Combining Mortality and Longitudinal Measures in Clinical Trials, *Statistics in Medicine*, 1999;18: 1341-1354.

4. SAS/STAT® 9.3 User's Guide, SAS Institute Inc., Cary, NC, 2011.

5. Dmitrienko, A. et. al., Analysis of Clinical Trials Using SAS: A Practical Guide, SAS Institute, Cary, NC, 2005.

6. Clopper C. J., Pearson E. S., The Use of the Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika, 1934, 26, 404-413.

7. Newcombe, R. G., Interval estimation for the difference between independent proportions: comparison of eleven methods, *Statistics in Medicine*, 1998, 17, 873-890.

8. Pocock, S.J., et.al., The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities, *European Heart Journal*, 2012, 33, 176-182.

9. Pocock, S.J., et. al., Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems, *Statistics in Medicine,* 2002, 21, 2917-2930.

10. Pocock, S.J., et. al., Regression to the mean in SYMPLICITY HTN-3, Implications for Design and Reporting of Future Trials, *JACC,* 2016, Vol. 68, No. 18, 2016-2025.

7 Acronyms and Abbreviations

Acronym/ Abbreviation	Term
6MWD	Six-Minute Walk Test Distance
6MWT	6 Minute Walk Test
A-G	Andersen-Gill
ASD	Atrial Septal Defect
AT	As-Treated
BNP	Brain Natriuretic Peptide
CDS	MitraClip Delivery System
CEC	Clinical Events Committee
CMS	Centers for Medicare and Medicaid Services
CPX	Cardiopulmonary Exercise Testing
DMC	Data Monitoring Committee
EF	Ejection Fraction
FDA	Food and Drug Administration
FMR	Functional Mitral Regurgitation
HF	Heart Failure
ITT	Intent To Treat
JFM	Joint Frailty Model
KCCQ	Kansas City Cardiomyopathy Questionnaire
LVEDD	Left Ventricular End Diastolic Dimension
LVEDV	Left Ventricular End Diastolic Volume
LVEF	Left Ventricular Ejection Fraction
LVESD	Left Ventricular End Systolic Dimension
LVESV	Left Ventricular End Systolic Volume
MI	Myocardial Infarction
MV	Mitral Valve
NT-pro BNP	N-terminal prohormone of Brain Natriuretic Peptide
NYHA	New York Heart Association
PP	Per Protocol

Acronym/ Abbreviation	Term
QoL	Quality of Life
RVSP	Right Ventricular Systolic Pressure
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SLDA	Single Leaflet Device Attachment
TEE	Transesophageal Echocardiogram

APPENDIX A Additional Analysis for CMS

A.1 Additional Clinical Outcomes for CMS

The following clinical outcomes requested by CMS will be reported for the Device and Control groups through 12 months follow-up post randomization, and compared between the Device and Control groups.

- Adjudicated stroke events at 12 months

- Adjudicated Transient ischemic attacks (TIAs) events at 12 months

- Major vascular complications (events) reported by sites at 12 months

  Major vascular complications will be reported through 12 months follow-up. Major vascular complications are defined based on the modified Valve Academic Research Consortium (VARC-2) definition as a composite of the following components:[1]

  a) Any aortic dissection, aortic rupture, cardiac rupture, cardiac valve rupture, ventricular rupture, arterial rupture, coronary artery perforation or cardiac perforation, or aneurysm/cardiac aneurysm/ cardiac or vascular pseudo-aneurysm
  b) Access site or access-related vascular injury (dissection, stenosis, perforation, rupture, arterio-venous fistula, vascular pseudoaneurysm, hematoma, nerve injury, compartment syndrome) resulting in serious adverse events
  c) Embolism (or embolus)/Thrombosis resulting in serious adverse events
  d) Any serious adverse event resulted in medical or surgical intervention to prevent permanent impairment to body structure or body function
  e) Any new lower extremity ischemia resulting in serious adverse events

- Renal complications reported by sites at 12 months

  Renal complications is a composite event of any acute kidney injury, renal failure, or renal insufficiency.

- Subsequent mitral valve operations (reoperations) at 12 months

  Subsequent mitral valve operations are defined to include any post "Treatment Visit" operations (open surgical or transcatheter) on the mitral valve in the Device group compared to the Control group.

[1] Kappetein AP et al. Updated Standardized Endpoint Definitions for Transcatheter Aortic Valve Implantation – The Valve Academic Research Consortium 2 Consensus Document. Journal of the American College of Cardiology, Vol 60, No. 15, 2012.

The number and proportion of subjects with the events through 12-month follow-up will be summarized by the binary method for both Device and Control groups. Subjects lost to follow-up or withdrawal before a time point (e.g. the 12-month visit) will be excluded from the summary unless they are known to have experienced the event prior to that time point (e.g. within 12 months).

A.2 CMS Subgroup Analysis

The following clinical outcomes will be analyzed by age subgroups (Age < 65 vs. Age ≥ 65) using Kaplan-Meier estimates. The treatment comparison in the age subgroup analysis is not powered for hypothesis testing and is not meant for confirmatory inference. Comparisons will be made between treatment arms within each age subgroup.

- Stroke at 12 months
- Major vascular complications (events) at 12 months
- Renal complications at 12 months
- All-cause mortality at 12 months