Statistical Analysis Plan

Cardiothoracic Surgical Trials Network

Evaluating the Benefit of Concurrent Tricuspid Valve Repair During Mitral Surgery (TR)

Sponsored by the National Heart, Lung, and Blood Institute (NHLBI) and the German Society for Thoracic and Cardiovascular Surgery (DGTHG)

CT Surgical Trials Network Research Group

Data and Clinical Coordinating Center
InCHOIR
Icahn School of Medicine at Mount Sinai
New York

Version: 1.0
Version Date: 7/14/2020

MODIFCATION HISTORY

Version Number	Date of Document Version	Significant Changes from Previous Authorized Version
1.0	7/14/2020	Not applicable – first version

TABLE OF CONTENTS

ABBREVIATIONS AND DEFINITIONS

AE	Adverse event
AV	Atrioventricular
CABG	Coronary artery bypass grafting
CEA	Cost-effectiveness analysis
CTSN	Cardiothoracic Surgical Trials Network
DCC	Data and Clinical Coordinating Center
DGTHG	German Society for Thoracic and Cardiovascular Surgery (Deutschen Gesellschaft für Thorax-, Herz- und Gefäßchirurgie)
DSMB	Data and Safety Monitoring Board
EAC	Event Adjudication Committee
EC	Ethics Committee
EQ-5D	EuroQoL
ICER	Incremental cost effectiveness ratio
InCHOIR	International Center for Health Outcomes & Innovation Research
IRB	Institutional Review Board
KCCQ	Kansas City Cardiomyopathy Questionnaire
LOS	Length of stay
MACCE	Major adverse cardiac and cerebrovascular event
MR	Mitral regurgitation
MV	Mitral valve
MVS	Mitral valve surgery
NHLBI	National Heart, Lung, and Blood Institute
NYHA	New York Heart Association
QALY	Quality adjusted life years
QOL	Quality of Life
REB	Research Ethics Board
RV	Right Ventricle
RVFAC	Right ventricular fractional area change
SAE	Serious adverse event
SAP	Statistical Analytical Plan
SF-12	Short Form 12
6MWT	Six Minute Walk Test
TAPSE	Tricuspid annular plane systolic excursion
TEE	Trans-esophageal echocardiography
TR	Tricuspid regurgitation
TTE	Transthoracic echocardiography
TV	Tricuspid valve

PURPOSE OF STATISTICAL ANALYTICAL PLAN (SAP)

The purpose of this SAP is to outline the planned analyses to be completed for the TR trial. The analyses identified in this SAP will be included in abstracts and manuscripts reporting the results of the trial. Exploratory analyses not necessarily identified in this SAP may also be performed. Any post hoc, or unplanned, analyses not explicitly identified in this SAP will be clearly identified as such in any published papers from this study. This SAP may be updated in response to additional developments, either within or outside the trial. All revisions will be made prior to the data lock and the primary analysis.

1. INTRODUCTION

The presence of moderate or severe TR is commonly encountered, affecting over 1.6 million people in the United States alone (Taramasso et al., 2012). As intrinsic pathology of the TV is rare, most TR is functional, defined as regurgitation in the presence of anatomically normal leaflets and chords. The precise mechanism by which functional TR develops is thought to be due to tricuspid annular dilation as well as by right ventricular (RV) enlargement and dysfunction associated with MV pathology in the presence of systolic and/or diastolic dysfunction and/or significant pulmonary hypertension (Dreyfus, Martin, Chan, Dulguerov, & Alexandrescu, 2015).

Although the clinical context in which TR occurs may influence prognosis, there are numerous reports that demonstrate the presence of TR being associated with increased mortality. Patients who do survive and develop more severe TR are more likely to develop New York Heart Association class III-IV symptoms (Groves, Lewis, Ikram, Maire, & Hall, 1991; Ruel et al., 2004) and decreased quality of life.

In 1967, Braunwald and colleagues demonstrated that correction of left-sided disease allowed for resolution of TR (Morrow, Oldham, Elkins, & Braunwald, 1967). In more recent years, however, this philosophy has been challenged by some, on account of observations that TR may in fact resolve only in a minority of cases. Overall data on the postoperative course and clinical sequelae of TR are conflicting, largely due to heterogeneous surgical management and MV pathologies.

In patients with severe TR already undergoing surgery for left valvular pathology, surgical correction is recommended by the AHA/ACC and ESC guidelines (Nishimura et al., 2014). Significant equipoise exists, however, as to the optimal approach for patients with only moderate TR or mild TR with annular dilation. Some argue that performing a TV annuloplasty at the time of MVS influences the incidence of right heart failure and improves long term survival, yet others believe the risk of an additional surgical procedure outweighs the potential benefit (Yilmaz et al., 2011).

The CTSN, therefore, designed a trial to evaluate the efficacy and safety of concomitant TR annuloplasty at the time of MVS. This document serves as the SAP for the TR trial.

2. STUDY OBJECTIVES

The overall objective of this study is to evaluate the safety and efficacy of tricuspid valve (TV) repair in the setting of mitral valve surgery (MVS) for degenerative mitral valve (MV) disease. Specifically, this study compares the surgical approach of combining TV annuloplasty with clinically indicated MVS to performing MVS alone.

2.1 Primary Objective

The primary aim of this trial is to evaluate the impact of these two surgical approaches on the composite endpoint of death and reoperation for tricuspid regurgitation (TR), or progression of TR, either by two grades from baseline (i.e. *prior* to randomization), *or* by the presence of

severe TR at 2 year follow-up.

2.2 Secondary Objectives

Secondary aims of this trial include assessment of the impact of these two surgical approaches on right heart performance and function, mortality, adverse events (AEs), quality of life (QOL), functional status, presence and severity of TR, and health resource use.

3. STUDY OVERVIEW

3.1 Study Design

This is a prospective, multi-center, randomized clinical trial in patients undergoing MVS for degenerative MV disease. Patients will be randomized (1:1) to receive either MVS + TV annuloplasty or MVS alone.

3.1.1 Study Duration and Time Points

All patients will be followed for 60 months post-randomization. Endpoints will be measured at 30 days, 6, 12, 18, and 24 months. Survival will continue to be measured after the 24-month visit via vital sign checks at 36, 48 and 60 months.

3.1.2 Randomization and Masking

Patients will be randomly assigned (1:1) to MVS + TV annuloplasty or MVS alone. Patient randomization will be stratified by TR severity and by clinical center. The randomization procedure will be performed intra-operatively, following the placement of the TEE probe and after visualization and confirmation of cardiac anatomy eligibility, in order to minimize the likelihood of enrolling patients in the study with unexpected surgical contra-indications to TV repair. Randomization will be performed centrally through a Web-based data collection system that automates the delivery of the randomization assignments. The treatment assignment will be viewed by the site coordinator electronically, in a secure fashion, and electronic verification of the treatment assignment will be required before proceeding with the treatment intervention.

Neither patients nor investigators will be blinded to treatment assignment due to the nature of the treatment intervention. Investigators will, however, be blinded to all data from other clinical sites, except serious unexpected AEs that are possibly or probably related to the investigational procedure for IRB/REB/EC reporting purposes. All echocardiograms will be analyzed by echocardiography core laboratory (Echo Core Lab) personnel who will be blinded to clinical outcomes. Adverse events (AEs) will be adjudicated by an Event Adjudication Committee (EAC) and trial oversight will be provided by an independent Data and Safety Monitoring Board (DSMB).

4. ANALYSIS POPULATIONS

Two populations will be used for all summaries and analyses.

Screened Population

The screened population will consist of all screened patients. A screened patient is defined as a consented subject who was referred to, or identified at a clinical site for consideration of entry into the study, and for whom some preliminary (i.e., medical record) data have been collected and/or reviewed.

Intent-to-Treat (ITT) Population

The ITT population will consist of all randomized subjects grouped by their assignment at randomization whether or not they actually received the treatment to which they were assigned. This sample will be used for summaries and analyses of the primary endpoint and the secondary clinical endpoints.

5. STUDY ENDPOINTS

5.1 Primary Endpoint

The primary endpoint will be a composite of death, reoperation for TR, or progression of TR from baseline, prior to randomization, by two grades or the presence of severe TR at 2 years post randomization.

Degree of TV regurgitation will be categorized according to American Society of Echocardiography guidelines as none/mild/moderate/severe. Trace regurgitation is also used in the event that regurgitation is barely detected.

5.2 Secondary Clinical Endpoints

The following secondary clinical endpoints will be assessed:

5.2.1 Clinical and Functional Outcomes

- A composite of major adverse cardiac and cerebrovascular events (MACCE), including stroke, death, and serious heart failure events by 24 months post-randomization
- Re-operation for TR by 24 months post-randomization
- NYHA classification at 30 days, 6, 12 and 24 months post-randomization.
- Diuretic Use at 30 days, 6, 12 and 24 months
- 6MWT at 30 days, 6, 12 and 24 months post-randomization.
- Gait Speed Test for Frailty at 12 and 24 months post-randomization.

5.2.2 Echocardiography

All echocardiography outcomes are measured by transthoracic 2D echocardiography unless otherwise noted

- Degree of TR at index hospital discharge
- Degree of TR at 12 and 24 months
- RV size at 12 and 24 months
- RV function (normal, mildly impaired, moderately impaired, severely impaired, peak tricuspid annular velocity, TAPSE, RVFAC) at 12 and 24 months
- Pulmonary artery pressure at 12 and 24 months
- RV volume at 12 and 24 months as measured by transthoracic 3D echocardiography

5.2.3 Quality of Life

- SF-12 at 6, 12, 18 and 24 months post-randomization.
- KCCQ at 6, 12, 18 and 24 months post-randomization.
- EQ-5D at 6, 12, 18 and 24 months post-randomization.

5.2.4 Survival

- Perioperative mortality (in-hospital or 30-day mortality)
- Mortality through 24 months post-randomization
- Mortality through 60 months post-randomization

5.2.5 Serious Adverse Events

- Frequency of SAEs
- AV-Block requiring pacemaker implantation
- New-onset atrial fibrillation

5.2.6 Hospitalizations

- Index hospitalization LOS and ICU days
- All-cause readmissions and readmissions for heart failure and TR re-operation through the first 30 days following surgery and through 24 months post-randomization

5.2.7 Economic Outcomes

- Cost
- Cost-effectiveness

6. STATISTICAL METHODOLOGY

6.1 General Principles

Study day will be calculated from the reference start date, and will be used to show the study days of assessments and events. Reference start date is defined as the date of randomization unless otherwise specified. In analyses of index length of stay, index ICU days, peri-operative (30 day) mortality, and peri-operative (30 day) readmissions the reference date is

the date of surgery. In the situation where the event date is partial or missing, study day, and any corresponding durations are to appear partial or missing in listings. If a missing event date, such as a discharge date for a hospital readmission, is necessary to calculate patient-time at risk, the missing event date will be imputed using the median length observed for similar events.

Continuous variables will be summarized using the following descriptive statistics: number of non-missing values, means, standard deviations, medians, interquartile range, maximum, and minimum. Categorical variables will be summarized using number of non-missing values, counts and percentages.

Rates of events will be calculated as the ratio of the total number of events recorded divided by the total patient-time. Total patient-time will be calculated by summing the time (in study time units, e.g., days, months or years) that patients were at risk for a specific event from the reference time point until either study exit or the end of the time period of interest. Rates and 95% confidence intervals will be reported.

Time-to-event variables will be summarized using the Kaplan-Meier method or, in the presence of competing risk, the Gray method (Gray, 1988).

For any variable measured at multiple points in time, change from baseline will be calculated as the difference between the value of the variable at a specific point in time (e.g. 1 year) minus the baseline value. Relative change from baseline will be calculated as the value of a parameter at a specific point in time minus the baseline value of the parameter divided by the baseline value of the parameter. Percent change will be calculated as the relative change multiplied by 100.

All hypothesis testing will be conducted at the 0.05 two-sided significance level unless otherwise specified. P-values will be rounded to three decimal places. P-values less than 0.001 will be reported as <0.001 in tables. P-values greater than 0.999 will be reported as >0.999.

Should any of the statistical methods proposed prove unsuitable during data analysis, more appropriate methods will be used. These include data transformation (for example to a logarithmic scale) to satisfy model assumptions such as normally distributed residuals with constant variance, the application of non-parametric techniques or the use of a different link function or modeling technique.

Additional ad-hoc analyses may be conducted as deemed appropriate.

All analyses will be conducted using SAS V9.4 or higher.

6.2 Missing Data

6.2.1 Missing Baseline Data

Missing baseline values will not be imputed and summaries will be based on all available

data.

6.2.2 Missing Primary Outcome Data

The plan for handling missing primary outcome data is outlined in section *6.5.1.3* below.

6.2.3 Missing Secondary Outcomes Data

In general, missing outcome values for secondary endpoints will not be imputed and analyses will be based on all available data. Multiple imputation may be used for specific analyses (e.g., cost analysis).

6.3 Crossover

Crossovers (patients who after randomization switch from the allocated treatment to the non-allocated treatment) are expected to be few in this trial. Patients randomized to TV annuloplasty who do not receive it during the trial can be considered crossovers. In addition, patients who are randomized to no annuloplasty but receive it during the index procedure are considered to have crossed over. As the primary analysis is by intention to treat, crossovers will be analyzed as belonging to the group to which they were randomized. Given the short duration between randomization and surgery, crossovers are assumed to be rare (no more than 3%).

6.4 Patient Characteristics

6.4.1 Patient Disposition

Disposition will be summarized in the screened and ITT populations.

Disposition summaries of the screened population will include:

- The number of patients screened
- The number and percentage of screened patients eligible
- The number and percentage of screened patients ineligible and the reasons for ineligibility summarized as the number and percentage of ineligible patients who met each ineligibility criteria
- The number and percentage of eligible patients randomized

Disposition in the ITT population will be summarized by randomization group and will include:

- The number of patients randomized
- The number and percentage of patients who received their assigned procedure
- The number and percentage of patients withdrawn or lost to follow-up by the primary outcome visit at 24 months and the primary reason for withdrawals
- The number and percentage of patients withdrawn or lost to follow-up by the final study visit at 60 months and the primary reason for withdrawals

6.4.2 Protocol Deviations

Protocol deviations and violations are defined as deviations from the procedures outlined in the protocol. There is no "Per Protocol" population defined for this study. All statistical analyses and summaries will be conducted on an intent-to-treat basis.

6.4.3 Patient Characteristics

6.4.3.1 Demographic characteristics

Demographics including age, gender, race and ethnicity will be summarized by randomization assignment using the appropriate descriptive statistics.

6.4.3.2 Baseline characteristics

Baseline characteristics will be summarized by randomization assignment using the appropriate descriptive statistics. The specific baseline variables collected are detailed in the protocol and include medical history, physical exam findings, medications, laboratory assessments, echocardiographic measures, quality of life, and functional status.

6.4.3.3 Operative characteristics

Operative data including primary procedure type, duration of operation, duration of aortic cross clamp time, duration of cardiopulmonary bypass time, and concomitant procedures will be summarized by randomization assignment using the appropriate descriptive statistics.

6.5 Primary and Secondary Outcome Analyses

All analyses will be performed using the ITT population.

6.5.1 Analysis of the Primary Outcome and Determination of Sample Size

The primary outcome is treatment failure defined as the composite of (1) death from any cause, (2) reoperation for TR, (3) presence of severe TR at two years post randomization or, for patients enrolled with less than moderate TR and annular dilatation, progression by two grades (i.e., from none/trace TR to moderate TR) at two years. The null hypothesis is that there is no difference in the probability of treatment failures at two years post randomization between patients randomized to undergo TV repair during MVS compared to patients randomized to undergo MVS alone. The primary null hypothesis will be tested in an intent-to-treat analysis using a 0.05 level two-tailed normal approximation (Wald) test.

6.5.1.1 Determination of Sample Size

Based on previously published data we assume that at two years post randomization, 25% of patients treated with only MVS will experience the primary composite endpoint (Goldstone et al., 2014; Koelling, Aaronson, Cody, Bach, & Armstrong, 2002; Nath, Foster, & Heidenreich, 2004). We believe a meaningful effect worth detecting is at least a 50% relative reduction to 12%, for patients undergoing TV annuloplasty in addition to MVS. A total of 400 patients, randomized with equal probability to each arm, provides approximately 90% power to detect such a

difference. For simplicity, power is based on a 0.05 level two-tailed chi-square test. The sample size takes account of a potential single interim analysis to be performed in addition to the final analysis, and a minimal (less than 3%) rate of crossover.

6.5.1.2 Primary Analysis

A log binomial regression model will be used to estimate and test differences in treatment failure between randomization groups. Similar to the logistic regression model, the log-binomial model is a generalized linear model. The models differ only in the link function used for the "success" probability p; logit (log odds) for logistic regression and log (log p) for log-binomial. The different links parameterize the model differently, with parameters of the log-binomial model yielding log relative risks rather than the log odds ratios of the logistic model.

The basic form of the log binomial models is:

$$\log P[Y_i = 1|X_{1i}, X_{2i}] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \,,$$

where Y_i is a binary indicator of treatment failure for the i^{th} patient, X_{1i} is a binary indicator of randomization assignment for the i^{th} patient and X_{2i} is an indicator of moderate or less TR at baseline, a factor by which randomization will be stratified. While randomization will also be stratified by randomizing center, the analysis will not adjust for center due to their relatively large number compared to the proposed sample size. The exponentiated estimate of β_1 ($e^{\widehat{\beta}_1}$) in this model is the risk ratio for the composite endpoint for patients randomized to TV repair compared to patients randomized to no TV repair. The risk ratio and its associated 95% confidence interval will be used to quantify the relative risk of the composite endpoint. Differences between randomization groups in the risk of the composite endpoint will be determined by testing the null hypothesis H_0:$\beta_1 = 0$ versus a two-sided alternative (H_1:$\beta_1 \neq 0$) using a 0.05 level intention-to-treat normal approximation test (i.e., the Wald test).

6.5.1.3 Imputation of Missing Primary Endpoint Data

We expect relatively few patients to be missing the primary endpoint due to withdrawal or refusal. Patients with missing data will have their 24 month status imputed via multiple imputation assuming that the data are missing at random. The imputation model will be stratified by randomization assignment and include age, sex, randomization strata for moderate or less TR at baseline, degree of TR at 6 months, and degree of TR at 12 months. Since this model includes a mixture of variables types (i.e. continuous, ordinal, and binary), a fully conditional specification method will be used (Berglund, Heeringa, & SAS Institute., 2014).

The main feature of the imputation approach is the creation of a set of clinically reasonable imputations for treatment failure for each patient with missing data. This will be accomplished using a set of repeated imputations created by predictive

models based on the majority of participants with complete data. The imputation models will reflect uncertainty in the modeling process and inherent variability in patient outcomes, as reflected in the complete data. Thirty datasets will be imputed.

After the imputations are completed, all of the data (complete and imputed) will be combined and the analysis performed for each imputed-and-completed dataset. Rubin's method of multiple (i.e., repeated) imputation will be used to estimate treatment effect (Rubin & Schenker, 1986).

6.5.1.4 Assessment of Balance of the Randomization

The success of the randomization procedure in balancing important covariates between randomization groups will be assessed. Continuous measures will be compared using t-tests, while chi-square tests will be used to compare categorical variables. As 400 patients will be randomized, no substantial imbalances are expected. However, should any covariate differ significantly between treatment groups at the 0.01 level, and be substantively large, we will adjust for those covariates in a secondary analysis of the primary endpoint.

6.5.1.5 Examination of Subgroups

A subgroup analysis of the primary endpoint will be performed with subgroups defined by whether or not patients received a CABG procedure during the initial surgery. This analysis will be performed in the same manner as that described for the primary analysis.

6.5.1.6 Impact of COVID-19 Pandemic

Due to the COVID-19 pandemic, patients randomized after January 2018 may be unable to do the on-site, in-person two-year follow up visit. To mitigate the risk of missing primary endpoint data, a revision to the protocol was issued via a study wide memo to widen the two-year visit window from ±60 days to ±120 days. The number and percentage of two-year visits missed due to COVID-19 and the number and percentage done outside of the original ±60 day window due to COVID-19 will be reported.

The primary analysis will be conducted as outlined in section 6.5.1.2 using all available two-year visit echocardiogram results with missing data imputed as outlined in section 6.5.1.3. A sensitivity analysis excluding data collected outside of the original ±60 day window due to COVID-19 will be conducted to explore the impact of enlarging the window of capture.

6.5.2 Analyses of Secondary Clinical Endpoints

6.5.2.1 Clinical and Functional Outcomes

MACCE: The rate of major cerebrovascular or cardiac events (defined as the

composite event of death, stroke, and serious heart failure events) will be compared between randomization groups over 24 months post-randomization using a Cox proportional hazards regression model.

Reoperation for TR: The difference in the rate of requiring subsequent TV annuloplasty after the initial MVS will be compared between randomization groups over 24 months post-randomization. Reoperation for TR may not occur because death from any cause precedes the event; thus, it is possible that censoring patients at all-cause mortality will lead to biased estimates when analyzing time to first event. Therefore, competing risks analysis using the methods of Fine and Gray (Fine & Gray, 1999) will be used to estimate group differences.

NYHA: The distribution of NYHA at 30 days, 6, 12 and 24 months will be presented for each randomization arm and compared using chi-squared tests.

Diuretic Use: The distribution of diuretic use at 30 days, 6, 12 and 24 months will be presented for each randomization arm and compared using chi-squared tests.

Six Minute Walk: Differences between groups in the distances travelled during the 6MWT at 30 days, 6, 12 and 24 months will be compared using Wilcoxon Rank-Sum tests.

Gait Speed Tests: Differences between groups in gait speed at 12 and 24 months will be compared using Wilcoxon Rank-Sum tests.

6.5.2.2 Echocardiography

Degree of TR: The distribution of the degree of TR at hospital discharge, 12, and 24 months will be presented for each randomization arm and compared using chi-squared tests. Between group differences in TR progression defined as presence of severe TR at 24 months, or for patients enrolled with less than moderate TR, progression by two grades compared to baseline will be compared using a two-tailed 0.05 level chi-squared test.

Additional echo parameters: RV size, RV function (normal, mildly impaired, moderately impaired, severely impaired, peak tricuspid annular velocity, TAPSE, RVFAC), and pulmonary artery pressure at 12 and 24 months will be compared between groups. RV volume as assessed by 3D TTE at 12 and 24 months will be compared between groups. Continuous variables will be compared for between group differences using the Wilcoxon Rank Sum test, and the level of function by chi-square test.

6.5.2.3 Quality of life

QOL will be measured using the KCCQ, SF-12, and EQ-5D. We will employ two

approaches to the analysis of QOL. The first will be to base the analysis on longitudinal mixed effects models. These models would predict outcome from treatment group and time. The mixed modeling approach requires an assumption that patient dropout is ignorable in that the probability of dropping out at any time is related only to previously observed data. Of course, this assumption may not hold, and moreover it is impossible to test robustly from the data at hand. An alternative approach, not subject to this criticism, will be to separate the data into strata defined by the time of death or dropout. We will then estimate a separate linear model, including a treatment effect, for the data in each stratum. This method, known as pattern-mixture modeling, is not sensitive to un-testable assumptions about the dropout mechanism because it models the data directly in strata defined by dropout time. The method of Wu and Bailey is an instance of pattern-mixture modeling (Wu & Bailey, 1989).

6.5.2.4 Survival

Perioperative Mortality: The distribution of vital status at 30 days post-surgery will be presented for each randomization arm and compared using chi-squared tests.

Mortality at 2 and 5 years post-randomization: Differences in the rate of mortality between randomization groups over the first 24 months and the planned 60 months of post-randomization follow-up will be assessed using Cox proportional hazards regression.

6.5.2.5 Serious Adverse Events

Differences in the rate of individual serious AEs, AV-Block requiring pacemaker implantation, and new-onset atrial fibrillation within 24 months post-randomization will be compared between randomization groups using Poisson regression. Ninety-five percent confidence intervals for the rate ratios for individual AEs for treatment with MVS and TV annuloplasty versus MVS alone will be computed.

6.5.2.6 Hospitalizations

Index hospitalization length of stay and days in Intensive Care: We will compare post-surgery hospital LOS and days spent in ICU between treatment groups, separately by region (North America and Germany). A Wilcoxon Rank-Sum test will be used to test for differences within each geographic subgroup.

Perioperative readmission by 30 days: The distribution of the incidence of all-cause readmission, cardiovascular readmission, heart failure readmission, and TR re-operation by 30 days post-surgery will be presented for each randomization arm and compared using chi-squared tests.

Hospital readmission by 2 years: Rates of all-cause hospitalizations and rates of

cardiovascular and heart failure specific hospitalizations, within two years of randomization will be compared using Poisson regression.

6.5.2.7 Costs and Cost-Effectiveness

Cost: Cost will be calculated by converting charges to cost using institution specific Ratio-of-Cost-to-Charges (RCCs). Institution-specific cost reports will be used to calculate RCCs for each major resource category. Cost data will only be collected in the North American sites. Costing data will be compared by Student's t test after log transformation. Independent predictors of cost, including baseline factors, operative factors and postoperative events, will be determined by multivariate regression analysis.

Cost-Effectiveness: The primary objective of the CEA is to estimate the incremental CE ratio (ICER) of the intervention under investigation as compared to the study-defined alternative. This ratio measures the ratio of the difference in costs and outcomes between the two study arms, with outcomes measured as quality-adjusted life-years (QALYs). QALYs reflect an individual's preference for both quantity and QOL in a single measure that facilitates comparisons across diverse treatment modalities. We will also compute net health benefits (NHB) as an alternative way of looking at cost-effectiveness. This parameter compares the incremental effectiveness of an intervention with the minimum health effect that society would demand in return for the investment; i.e., with the health produced by investing at the societal ceiling cost-effectiveness ratio (CR).

Costs will be estimated as discounted incremental health care costs, and effectiveness will be measured as the discounted increment in quality-adjusted life years. A secondary objective will be to identify disease- and patient-related factors that predict high costs of care following the intervention. All CE ratios will be reported with probability intervals to reflect the level of uncertainty in the clinical estimates used in the model and the underlying economic assumptions. We anticipate that the distribution of costs will be skewed to the right. If this violates the assumption of normality, we will modify the method using the nonparametric Bayesian bootstrap. We will use standard discount rates for both QALYs and costs.

We will calculate the ICER based on actual trial data and also develop a model to project long-term cost-effectiveness. Sensitivity analyses will be performed to estimate several sources of uncertainty, including sampling variation and variations in discount rates.

6.6 Interim Analysis

We plan to perform a single interim analysis with respect to the primary endpoint to give the option of stopping early should results strongly favor one arm or the other. The proposed timing of this analysis is at 0.5 on the information scale, i.e., after one-half of patients (200) reach the primary endpoint. The utility of performing this analysis will depend on the rate of

accrual of patients into the trial. As the decision to terminate early would likely occur after most, if not all, patients were randomized, the principal benefit of early termination would be prompt dissemination of results, and no further randomization to an inferior treatment. A group sequential procedure will be used to allow for flexibility in the number and timing of interim analyses should the DSMB choose to modify the proposed plan, or should accrual mitigate the usefulness of an interim look. We will use the Lan-DeMets approach, implementing an O'Brien-Fleming-type spending function that allots most of the type I error to the final look. The resulting critical values to be used for each analysis are 2.963 at the first interim analysis, 1.969 at the final analysis.

In addition to the ethical concern of continuing a trial that shows a clear benefit in favor of one treatment, there is also a corresponding ethical concern of continuing a trial that has little chance of ever showing a benefit of one treatment compared to the other. We propose that the trial's conditional power, under the original alternative hypothesis, be computed at the interim look and that the DSMB use this to determine whether randomization, if not completed, be halted for futility. We propose that consideration be given to halting the trial for futility if, given the data up to the point of the interim analysis, the probability of detecting a relative 52% reduction (from 25% to 12%) in the incidence of treatment failure for patients receiving TV annuloplasty in addition to MVS and patients randomized to MVS alone is less than 10%.

We do not propose any a priori stopping criteria based on AEs. The treatments in this trial are not experimental, and have well known AE profiles. Moreover, we believe that incident rates of AEs and mortality must be interpreted along with information about the consistency of related measures, consistency across centers, data completeness, and any external factors including scientific developments that might impact patient safety. In addition to considering the data generated by this trial, the DSMB will consider all relevant background knowledge about the treatment of MR. The DSMB would be capable, and uniquely suited, to determine decisions for convening outside the schedule of meetings, and to determine decisions to suspend or terminate the trial. These decisions should be at the discretion of the DSMB alone, based on all relevant information reported by the DCC and the Medical Monitors. We therefore recommend that the DSMB should be responsible for defining its deliberative processes, including event triggers that would call for an unscheduled review.

REFERENCES

Berglund, P., Heeringa, S., & SAS Institute. (2014). *Multiple imputation of missing data using SAS*. Cary, N.C.: SAS Institute.

Dreyfus, G. D., Martin, R. P., Chan, K. M. J., Dulguerov, F., & Alexandrescu, C. (2015). Functional Tricuspid Regurgitation A Need to Revise Our Understanding. *Journal of the American College of Cardiology, 65*(21), 2331-2336. doi:10.1016/j.jacc.2015.04.011

Fine, J., & Gray, R. (1999). A proportional hazards model for the sub distribution of a competing risk. *J Am Stat Assoc.*(94), 496–509.

Goldstone, A. B., Howard, J. L., Cohen, J. E., MacArthur, J. W., Jr., Atluri, P., Kirkpatrick, J. N., & Woo, Y. J. (2014). Natural history of coexistent tricuspid regurgitation in patients with degenerative mitral valve disease: implications for future guidelines. *J Thorac Cardiovasc Surg, 148*(6), 2802-2809. doi:10.1016/j.jtcvs.2014.08.001

Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Annals of Statistics, 16*(3), 1141-1154. Retrieved from <Go to ISI>://WOS:A1988Q161700019

Groves, P. H., Lewis, N. P., Ikram, S., Maire, R., & Hall, R. J. (1991). Reduced exercise capacity in patients with tricuspid regurgitation after successful mitral valve replacement for rheumatic mitral valve disease. *Br Heart J, 66*(4), 295-301. doi:10.1136/hrt.66.4.295

Koelling, T. M., Aaronson, K. D., Cody, R. J., Bach, D. S., & Armstrong, W. F. (2002). Prognostic significance of mitral regurgitation and tricuspid regurgitation in patients with left ventricular systolic dysfunction. *Am Heart J, 144*(3), 524-529. doi:10.1067/mhj.2002.123575

Morrow, A. G., Oldham, H. N., Elkins, R. S., & Braunwald, E. (1967). Prosthetic replacement of the mitral valve. Preoperative and postoperative clinical and hemodynamic assessments in 100 patients. *Circulation, 35*(5), 962-979. doi:10.1161/01.cir.35.5.962

Nath, J., Foster, E., & Heidenreich, P. A. (2004). Impact of tricuspid regurgitation on long-term survival. *J Am Coll Cardiol, 43*(3), 405-409. doi:10.1016/j.jacc.2003.09.036

Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., 3rd, Guyton, R. A., . . . Members, A. A. T. F. (2014). 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation, 129*(23), 2440-2492. doi:10.1161/CIR.0000000000000029

Rubin, D. B., & Schenker, N. (1986). Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association, 81*(394), 366-374. Retrieved from <Go to ISI>://WOS:A1986C648000012

Ruel, M., Rubens, F. D., Masters, R. G., Pipe, A. L., Bedard, P., & Mesana, T. G. (2004). Late incidence and predictors of persistent or recurrent heart failure in patients with mitral prosthetic valves. *J Thorac Cardiovasc Surg, 128*(2), 278-283. doi:10.1016/j.jtcvs.2003.11.048

Taramasso, M., Vanermen, H., Maisano, F., Guidotti, A., La Canna, G., & Alfieri, O. (2012). The growing clinical importance of secondary tricuspid regurgitation. *J Am Coll Cardiol, 59*(8), 703-710. doi:10.1016/j.jacc.2011.09.069

Wu, M. C., & Bailey, K. R. (1989). Estimation and Comparison of Changes in the Presence of Informative Right Censoring - Conditional Linear-Model. *Biometrics, 45*(3), 939-955. doi:Doi 10.2307/2531694

Yilmaz, O., Suri, R. M., Dearani, J. A., Sundt, T. M., Daly, R. C., Burkhart, H. M., . . . Schaff, H. V. (2011). Functional tricuspid regurgitation at the time of mitral valve repair for degenerative leaflet prolapse: The case for a selective approach. *Journal of Thoracic and Cardiovascular Surgery, 142*(3), 608-613. doi:10.1016/j.jtcvs.2010.10.042