BeAT-HF Extended Phase Morbidity and Mortality

Statistical Analysis Plan Supplement

December 21, 2022

Confidential

Document Number: 360050-003 Rev. E v1.0
NCT Number: NCT02627196

CVRx®

BeAT-HF Extended Phase Morbidity and Mortality

Statistical Analysis Plan Supplement

December 21, 2022

Confidential

Document Number: 360050-003 Rev. E v1.0

TABLE OF CONTENTS

TABLE OF TABLES

1. BACKGROUND

This Morbidity and Mortality Statistical Analysis Plan (SAP) supplement accompanies the main Statistical Analysis Plan (360050-001). Where applicable, the analyses defined in this Morbidity and Mortality SAP supplement supersedes the main SAP and the current Bayesian Adaptive Design Report (BADR). In particular; this SAP supplement describes the data analyses planned to demonstrate device benefit for the morbidity and mortality endpoint to support the M&M supplemental PMA (PMA-S) Submission.

After evaluating the pre-planned Expedited Phase data as defined in the main SAP, a large, important and clinically relevant population emerged. This population subgroup is characterized by having an NT-proBNP < 1600 pg/ml. This subgroup is the focus of this SAP supplement.

The M&M analyses will support a PMA-S submission for the following Indication for Use statement: The BAROSTIM NEO System is indicated for patients with heart failure. Heart failure is defined as: New York Heart Association (NYHA) functional Class III or Class II (who had a recent history of Class III), a left ventricular ejection fraction (LVEF) ≤ 35% and a NT-proBNP < 1600 pg/ml despite being treated with the appropriate heart failure guideline-directed medical therapy, excluding subjects eligible for Cardiac Resynchronization Therapy (CRT) according to AHA/ACC/ESC guidelines.

In Table 1 below, data that has already been analyzed (data as of 03OCT2018) in the Interim Data for the respective endpoint is called *interim data*. The prospective confirmatory data, which has yet to be accrued and/or analyzed for the respective endpoint, is called the *prospective data*. The **Subgroup** refers to subjects with a baseline NT-proBNP < 1600 pg/ml. This SAP describes the plans for analyzing the Final M&M data to support the M&M PMA-S Submission (last column in Table 1 below).

Table 1: BeAT-HF PMA Submissions and Data Elements

Data	Expedited PMA Submission Interim Dec 2018 / Prospective Apr 2019			PMA-S Submission*
	Interim Data (Submitted Dec 19, 2018)		Prospective Confirmatory Data	M&M
	All	Subgroup	Subgroup	Subgroup
6-Month				
MANCE	Interim data	Interim data	Interim + prospective data[p]	
6MHW	Interim data	Interim data	Prospective data	
QOL	Interim data	Interim data	Prospective data	
NT-proBNP	Interim data	Interim data	Interim + prospective data[b]	
M&M trending	Interim data		Interim + prospective data[p]	
Final M&M				Interim / prospective data[p]

*Future M&M SAP supplement; [p] pooled; [b] borrowing

This SAP includes sample size justification for randomizing an additional 72 subjects prospectively that will contribute to an increased sample size for the final M&M data. Summary details on the M&M endpoint analysis, analysis populations and sample sizes can be found in Table 2 below.

Updated 22Dec2022: Due to COVID-19, randomizations were completed on July 29. 2020 with 59 additional prospective subjects randomized (13 less than the originally planned 72 additional subjects). Thus, the total analysis population for the M&M endpoint is 319 randomized subjects instead of 331 noted in Table 2 below. Note: Four were randomized to treatment but were never implanted, so the total randomized is n=323. Table 2 provides the original planned sample size numbers.

Table 2: Original Endpoints Details (Updated 22Dec2022)

Evidence	Endpoint	Hypothesis	Analysis Population
Final M&M	Rate of morbidity and mortality events	Relative rate of events in treatment arm superior compared to rate of events in control arm Success criteria: Z-value assessing treatment effect at pre-specified interim analysis greater than Hwang-Shih-DeCani group-sequential upper bound Operating characteristics: type I error = 0.044 (one-sided) power = 0.860 Minimum observed relative reduction required: 28% (pooled data)	**Prospective data:** Subjects with • NT-proBNP < 1600 at baseline • Potential additional follow-up after 03OCT2018 Maximum sample size contributing additional data: N=241 subjects randomized prior to interim still under follow-up + 72 additional subjects to be randomized **Interim data:** Subjects with • NT-proBNP < 1600 at baseline • Follow-up through 03OCT2018 Sample size N=249, Treatment=119, Control=130 **All data:** • NT-proBNP < 1600 at baseline • Any post-activation follow-up N = 331 (Updated n=319) (241 interim & prospective, 72 prospective only (Updated n=59), 18 interim only)

2. MORBIDITY AND MORTALITY (M&M) ENDPOINT DETAILS

This section provides details for endpoints that contain morbidity and mortality data, particularly the cardiovascular (CV) mortality and heart failure (HF) morbidity endpoints. **All morbidity and mortality events are defined in the BeAT-HF Clinical Events Committee (CEC) Charter**.

To comprehensively measure the impact of Barostim therapy on mortality and heart failure morbidity, several definitions of heart failure morbidity will be evaluated. The inclusion of several definitions is necessary to evaluate the impact of COVID-19 on mortality, treatment for heart failure (both inside and outside the hospital/ER), and challenges with obtaining morbidity and mortality documentation from the institutions. Relevant Sensitivity Analysis (Section 4) and Supporting Morbidity and Mortality Analyses (Section 5) may be performed for each of these definitions, if appropriate.

The primary morbidity and mortality endpoint will be analyzed when a minimum of 320 events have been adjudicated. The events to be included in the M&M endpoint analyses are defined as follows:

- After the Clinical Event Committee (CEC) meeting in which the adjudication of the 320nd event occurs (CEC 320 Meeting), the ENDPOINT CLOSING DATE is established. This ENDPOINT CLOSING DATE will be the latest hosp/ER admission or censoring event onset date across ALL events adjudicated up to and including all events adjudicated during the CEC 320 Meeting (may exceed 320).
- Events that occur before the ENDPOINT CLOSING DATE will be included in the M&M endpoint analyses.

The assumptions for the primary morbidity and mortality endpoint are described in Section 7.

3. PRIMARY AND FOCUSED ENDPOINTS

3.1. Primary Endpoint: Rate of CV Mortality and HF Morbidity

3.1.1. Rate of Cardiovascular Mortality and Heart Failure Morbidity

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management, reduces the rate of cardiovascular mortality or worsening heart failure that leads to hospitalization or emergency room visit, cardiac assist device or heart transplant.

Subjects will be censored at the time of a terminal event (i.e., death, permanent cardiac assist device, heart transplant, or novel implantation of a device to receive Cardiac Contractility Modulation) or end of follow-up, whichever occurs first.

The end of follow-up for analysis of morbidity and mortality will occur on the last date where adverse event status is known for the subject. This definition of follow-up time ensures that only time periods where morbidity and mortality status has

been assessed are included. End of follow-up is defined for each subject as the later date of:

- Last study visit
- Last date of adverse events reported in the database:
  - If last adverse event is a hospitalization, use discharge date
  - If last event is not a hospitalization, use resolution date if resolved, else use onset date
- Date subject completed the final, protocol defined end of study visit (note: this only occurs at end of study)

This end of follow-up date will be used to calculate follow-up time for the negative binomial model.

3.1.2. Statistical Hypothesis Test

The rate of endpoint events will be compared between the two groups using a negative binomial model. The model will be adjusted for the number of heart failure hospitalizations within 12 months prior to enrollment to adjust for any potential imbalance between the two arms.

The statistical hypothesis test for the primary endpoint, stated in terms of a null and alternative hypothesis is:

Ho: $\beta_1 = 0$
Ha: $\beta_1 < 0$

where β1 is the natural log of the ratio of events per patient-year between groups (treatment/control). The treatment effect of the device as a relative reduction in the endpoint event rate is estimated by 1-exp(β1). The parameter β1 is referred to as θ in the BADR where the likelihood function for the joint negative binomial model is described.

3.1.3. Group sequential design

Morbidity and mortality will be evaluated using a frequentist group sequential design. The group sequential boundaries are specified by the Hwang-Shih-DeCani spending function with $\gamma = -3$. The parameter γ indexes the Hwang-Shih-DeCani spending function such that $\gamma = -4$ is similar to an O'Brien-Fleming boundary and $\gamma = 1$ is similar to a Pocock boundary. Analyses are planned based on data accumulated through the time at which 140, 180, 240, and 320 primary endpoint events have occurred in the trial. These three interim analyses and a final analysis after 320 events have occurred are planned due to uncertainty regarding the treatment effect size. A large treatment effect would result in stopping the study at one of the interim analyses while a more moderate treatment effect would require more events in order to conclude that the treatment effect is statistically significant.

The boundaries of the Hwang-Shih-DeCani spending function at each interim analysis will be determined based on the estimated information fraction at the time of analysis and at any previous interim analyses. The information fraction for a negative binomial model estimate of the treatment effect comparing two randomized arms is defined in Mütze et al. (2018)[1]. The information fraction depends on the event rate in each arm, the length of follow-up of each patient, and the variance (dispersion) of the negative binomial distribution. A consistent estimate of the information can be obtained by plugging in maximum likelihood estimates from the interim data. At each interim analysis, the total information of the trial at 320 events will be estimated based on simulation through 320 events using the negative binomial model parameter estimates (e.g., dispersion) obtained at the time of interim analysis. The estimated information at the time of the interim analysis divided by the simulated total information at 320 events will provide the information fraction at the interim analysis to determine the spending function boundary.

3.1.4. Analysis Method

The hypothesis test will be based on a Z-statistic for the estimate of β_1 from a negative binomial regression model. Although randomization is expected to result in balanced baseline covariates between the two arms, the model will include the baseline covariate of number of heart failure hospitalizations 12 months prior to enrollment to adjust for any potential imbalances between the two arms. In SAS (version 9.4 code) this model is fit as follows:

```
proc genmod;
model nev=trt hfh / dist=negbin offset=lfutime;
run;
```

where nev is the number of events, trt is an indicator variable of randomization to the treatment arm, hfh is the number of heart failure hospitalizations within the 12 months prior to enrollment, and lfutime is the logarithm of the follow-up time. The Z-statistic (Z) is the parameter estimate for trt divided by the associated standard error. A Wald test one-sided nominal p-value is calculated by calculating the standard normal tail probability $\Phi(Z)$. This simplified model has previously been demonstrated[2] to provide treatment effect estimates in strict agreement with those obtained with the more complex joint model defined in the BADR.

Success for demonstrating a reduction in morbidity and mortality is determined by the z-statistic crossing the group sequential boundary at one of the pre-specified analyses whose timing is determined by reaching pre-specified numbers of events.

3.1.5. Analysis Population

The analysis population for the prospective data to be analyzed for this endpoint is all randomized subjects with a baseline NT-proBNP < 1600 that have post-activation follow-up. Endpoint follow-up evaluation for the morbidity and

mortality endpoint begins at the actual device implant (device subjects) or anticipated implant date that is recorded prior to randomization (medical management subjects). All events adjudicated by the CEC as meeting the endpoint event definitions will be included in the analysis. Both prospective and previously analyzed follow-up time periods and events will be included in the analysis.

3.2. Focused Endpoints

Based on discussions in November 2022 between the FDA and the BeAT-HF Steering Committee, as well as CVRx and an independent statistician, a set of focused and clinically relevant endpoints were pre-specified. The purpose of these pre-specified focused endpoints is to provide a structure to evaluate the totality of the data while limiting the potential for a type I error inflation among non-primary endpoints. The primary endpoint and analysis method remains unchanged, and the analysis of all focused endpoints will be performed independently from the results of the primary endpoint.

The pre-specified focused endpoints were identified prior to the unblinding of the data. These focused endpoints are a subset of the endpoints and analyses pre-specified in the Statistical Analysis Plan, with minor clinically justified and statistical clarifications. The analyses of the focused endpoints will be based on descriptive statistics, nominal two-sided 95% confidence intervals and p-values. No statistical adjustment for multiple comparisons will be made.

These analyses are intended to support any updated labeling for patients with heart failure.

3.2.1. Rate of CV Mortality and HF Morbidity (Expanded Definition)

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management, reduces the rate of cardiovascular mortality or worsening heart failure that leads to hospitalization (expanded definition) or emergency room visit (expanded definition), cardiac assist device or heart transplant.

As this endpoint may include additional events due to the expanded definition, and the planned censoring (see Section 2.1), statistical hypothesis (see Section 2.1.2), analysis method (see Section 2.1.4) and population (see Section 2.1.5) is the same as the primary endpoint, the endpoint should have at least the power of the primary endpoint.

The expanded definition is defined in the CEC Charter and includes hospitalizations that are determined to be heart failure hospitalizations by a CEC consensus yet may not have all of the documentation available.

Note: This analysis may be updated to include all events that are adjudicated up to the final Clinical Events Committee meeting, regardless of ENDPOINT CLOSING DATE defined in Section 2.

Rationale: This endpoint is consistent with the primary endpoint except that it includes events that are hospitalizations / ER visits that are determined to be heart failure hospitalizations by a CEC consensus yet may not have all the documentation available. This enables events that are of equal significant to the primary endpoint events to be included in evaluating the M&M benefit.

3.2.2. Win Ratio Expanded – MLWHF QoL

A Win Ratio analysis[4] will be performed using the components of the morbidity and mortality endpoints as well as Minnesota Living with Heart Failure Quality of Life at 12 months. The following hierarchy will be used:

1. CV Death
2. Heart transplant or LVAD
3. Number of hospitalizations or emergency department visits for heart failure using the expanded definition of heart failure
4. Number of unscheduled clinic visits w/IV diuretic using the expanded definition for heart failure
5. Change from baseline in Minnesota Living with Heart Failure Quality of Life at 12 months (5 points is minimal clinically relevant difference).

The Win Ratio analysis will use all pairs of subjects between the two study arms to determine the number of wins and losses for the treatment versus control arm.

For each pair, follow-up is defined by the shorter of the follow-up periods. For example, if one patient has 1 year of follow-up and the 2nd has 2 years, the follow-up time used for comparison is 1 year. If the first subject in a pair is censored prior to an event time for the second subject in a pair, that pair is a tie. This produces the shared duration of follow-up for each pair, avoiding bias due to differential follow-up.

Any missing data for MLWHF QOL at 12M will be handled with single imputation via fully conditional specification linear regression based on baseline characteristics (MLWHF QOL, age, sex, LVEF, NYHA, NT-proBNP) and 6M MLWHF QOL. Single imputation is sufficient for the win ratio analysis, as the goal of the imputation is to produce an individual value for a missing pair to facilitate breaking ties.

The p-value comparing the two arms will be calculated using the Schoenfeld-Finkelstein method. A 95% confidence interval for the Win Ratio will be calculated using the bootstrap method.

Rationale: The Win Ratio incorporates both clinical events as well as patient-centric quality of life, providing a comprehensive evaluation of the impact of the therapy for patients. These events are weighted based on severity, where cardiovascular death is considered a more serious event than other non-fatal events such as LVAD / heart transplant, which is followed by heart failure hospitalizations. The last comparison is Minnesota Living with Heart Failure

Quality of Life ensuring that a clinically meaningful improvement in quality of life is evaluated.

3.2.3. Days in Hospital for Heart Failure

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management, decreases the number of days in the hospital for heart failure. Analysis will be based on a similar negative binomial model as the primary endpoint, including the definition of follow-up time and the same negative binomial regression model for estimation and producing a nominal two-sided 95% confidence interval and p-value for the treatment effect. Similar to the primary endpoint, to further incorporate the impact of the timing that the events occur over follow-up, alternative methods of analysis will be performed as described in Section 5.3.

Rationale: Days in the hospital for heart failure represents a patient-relevant outcome that measures the severity of their heart failure as measured by hospitalization days. The endpoint accounts not just for the number of heart failure hospitalizations, but also the length or severity of those hospitalizations. The days in hospital for heart failure can provide relevant information for patient and physician decision-making.

3.2.4. Long Term Sustainable Improvement

Six Minute Hall Walk (6MHW)

The mean improvement in 6MHW between baseline and 12 months by treatment arm will be presented with means and confidence intervals using a regression model of change from baseline with covariates for randomized group and baseline value. A generalized estimating equation (GEE) repeated measures model with a compound symmetry working correlation will be used to account for within subject correlation over time and to estimate the difference between the randomized arms over all times at which the endpoint is collected, adjusted for baseline.

Rationale: Improving symptoms is important to patients, particularly an improvement in exercise capacity (6MHW). Previously, a clinically and statistically significant symptomatic improvement was established at six months for 6MHW. This endpoint will evaluate the long term, sustainability of an improvement in 6MHW between the treatment arms.

Minnesota Living With Heart Failure (MLWHF)

The mean improvement in MLWHF between baseline and 12 months by treatment arm will be presented with means and confidence intervals using a regression model of change from baseline with covariates for randomization group and baseline value. A generalized estimating equation (GEE) repeated measures model with a compound symmetry working correlation will be used to account for within

subject correlation over time and to estimate the difference between the randomized arms over 12 months of follow-up, adjusted for baseline.

Rationale: Improving symptoms is important to patients, particularly an improvement in Quality of Life. Previously, a clinically and statistically significant symptomatic improvement was established at six months for QOL. This endpoint will evaluate the long term, sustainability of an improvement in QoL between the treatment arms.

NT-proBNP

The mean improvement in NT-proBNP between baseline and 12 months by treatment arm will be presented with means and confidence intervals using a regression model on the log transformed NT proBNP change from baseline with covariates for randomization group and baseline value. A generalized estimating equation (GEE) repeated measures model with a compound symmetry working correlation will be used to account for within subject correlation over time and to estimate the difference between the randomized arms over all times at which the endpoint is collected, adjusted for baseline.

Rationale: An improvement in NT-proBNP is associated with a decreased risk of morbidity and mortality events. Previously, a clinically and statistically significant relative reduction in NT-proBNP was established at six months. This analysis will evaluate the long term, up to 12 months, sustainability of a relative reduction in NT-proBNP between the treatment arms.

4. SENSITIVITY ANALYSES FOR MORBIDITY AND MORTALITY ENDPOINTS

The morbidity and mortality sensitivity analysis defined in this SAP includes, and supersedes, the morbidity and mortality sensitivity analysis defined in the main Statistical Analysis Plan (360050-001). These sensitivity analyses will be performed on the primary endpoint, and may also be performed, as appropriate, on the secondary and ancillary endpoints.

4.1. Sensitivity Analyses Due to Additional Therapeutic Interventions

Progression of heart failure may result in additional therapeutic intervention, as defined in Table 3 below, leading to potential bias in the assessment of between arm differences. In order to assess the potential impact of adoption of more intensive therapeutic intervention on the estimation of treatment effect for the morbidity and mortality endpoint, sensitivity analyses will be conducted. These sensitivity analyses acknowledge the potential bias introduced by use of a more intensive therapeutic intervention and serve to assess the degree of bias.

Table 3: Additional Therapeutic Intervention Sensitivity Analyses

#	Therapeutic Intervention	Definition
1	New medication class	Intervention is defined as addition of a new class of medication (e.g., ARNI or ivabridine) not part of a subject's baseline medication regimen. ARNI and ivabridine are new medications becoming available for the treatment of heart failure patients at the time of study initiation. Increasing adoption of these medications, or other novel classes of medication yet to be developed, has the potential to impact estimation of the device treatment effect relative to medical management.
2	Medication dose increase	A qualifying medication increase is defined as a 100% increase in the effective dose of medications within a specific drug class (e.g., ACE inhibitors, beta-blockers, etc.) compared to baseline dose for any class of medication that is part of a subject's baseline medication regimen.
3	Cardiac resynchronization therapy (CRT)	Intervention is defined as the addition of a CRT LV lead or the implantation of a CRT device.
4	CardioMEMS device	Intervention is defined as the implantation of a CardioMEMS device. An additional analysis will be performed based on first intervention due to CardioMEMS usage if information available.
5	BAROSTIM NEO device	Intervention is defined as non-protocol implantation of a BAROSTIM NEO device
6	Any intervention (1-5 above)	Overall assessment of sensitivity to any of the specified interventions.
7	Medication class or dose (1-2 above)	Overall assessment of sensitivity to specified medication changes.
8	Device interventions (3-5 above)	Overall assessment of sensitivity to specified device interventions.

The sensitivity analyses defined below will be performed for each of the additional therapeutic interventions in Table 3:

- An analysis will be performed with follow-up censored at the earliest time that a specified therapeutic intervention occurs. For this sensitivity analyses, the earliest occurrence of the specified therapeutic intervention will NOT be considered an event contributing toward the endpoint.
- An analysis will be performed that treats a specified therapeutic intervention as a terminal event (i.e., no other follow-up time or events past the therapeutic intervention will count). For this sensitivity analyses, these terminal events will be considered as an event contributing toward the endpoint.
- An analysis that counts all specified therapeutic interventions as events contributing toward the endpoint, with follow-up time accruing until last available contact/death.
- An analysis imputing events for the post-therapeutic intervention period, separately for the randomized arms. Imputation will be based on subject's baseline covariates and pre-therapeutic intervention data and the amount of follow-up time available post-therapeutic intervention. For this sensitivity analyses, these imputed events will be considered as an event contributing toward the endpoint.

4.2. Sensitivity to imbalance in baseline covariates

Randomization is important to ensure that the treatment arms are balanced at baseline. Although subjects in the Prospective Confirmatory Data Cohort were randomized, this randomization took place prior to the identification of the baseline NT-proBNP subgroup. As such, it may be possible that there are imbalances in the treatment arms with respect to baseline variables in the subgroup. To mitigate the potential impact of any imbalances, a propensity score analysis will be performed to adjust for observed imbalances in the treatment arms with respect to observed baseline characteristics.

Propensity scores will be estimated using logistic regression with randomized arm as the dependent variable and a pre-specified list of baseline characteristics that may be related to variation in endpoint data as independent variables. Propensity scores will be used in an inverse probability of treatment weighted (IPTW) model, such that all patient data will be included in the analysis. As weights may be unstable for patients with a low or high propensity score, a maximum weight will be specified. The analysis will use average treatment effect in treated subjects (ATT) weights for the weighted analysis and standardized differences to assess the degree of imbalance between baseline variables. ATT weights and standardized differences have been defined in Austin (2011)[3]. Baseline variables will be considered suitably balanced if the standardized difference is less than 0.1. For each endpoint, propensity score weights will be used to perform a weighted version of the primary regression analysis defined for that endpoint in order to estimate a propensity-score adjusted treatment effect.

The following baseline covariates will be included in the IPTW: Gender, age (<65 year vs >=65 years), race, NT-proBNP, 6MHW, history of heart failure hospitalization, history of either CABG, MI or atrial fibrillation.

4.3. Intention-to-Treat Analysis

To assess the potential for bias in the comparison of the treatment arms due to differences in withdrawal during follow-up, a sensitivity analysis will be performed on all randomized subjects in which follow-up time will begin at time of randomization, instead of "activation" [defined as actual device implant (device subjects) or anticipated implant date that is recorded prior to randomization (medical management subjects)]. All events that occur from time of randomization will be included in the analysis, even the events in subjects that have no post-activation follow-up.

4.4. Per Protocol Analysis

A sensitivity analysis will be performed with events and follow-up censored at the first time a subject receives therapy out of compliance with the protocol. For device arm subjects, the date that device therapy is turned off will be the censoring date. For medical management arm subjects, if crossover occurs, the date that device therapy is turned on will be the censoring date.

5. SUPPORTING MORBIDITY AND MORTALITY ANALYSES

These endpoints will provide additional supporting analyses to aid interpretation of the morbidity and mortality data. These analyses may include events included in the heart failure expanded definition and/or heart failure equivalent events, as defined in Section 2, as appropriate for the endpoint.

The analyses of the focused endpoints will be based on descriptive statistics, nominal two-sided 95% confidence intervals and p-values. No statistical adjustment for multiple comparisons will be made.

These analyses are intended to support any updated labeling for patients with heart failure.

5.1. Cardiovascular Mortality and Heart Failure Morbidity N

These two endpoints are similar to the primary and first focused endpoint, except they also include the following hospitalization equivalent events: Hospital-based observation unit and unscheduled clinic visits.

5.1.1. Rate of CV Mortality and HF Hospitalization / Hospitalization Equivalent

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management, reduces the rate of cardiovascular mortality or worsening heart failure that leads to any of the following events:

- Hospitalization for heart failure OR
- Emergency room visit for heart failure OR
- Hospital-based observation unit for heart failure OR
- Unscheduled clinic visit for heart failure OR
- Cardiac assist device OR
- Heart transplant

5.1.2. Rate of CV Mortality and HF Hospitalization / Hospitalization Equivalent (Expanded Definition)

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management, reduces the rate of cardiovascular mortality or worsening heart failure that leads to any of the following events:

- Hospitalization for heart failure (expanded definition) OR
- Emergency room visit for heart failure (expanded definition) OR
- Hospital-based observation unit for heart failure (expanded definition) OR
- Unscheduled clinic visit for heart failure (expanded definition) OR
- Cardiac assist device OR

- Heart transplant

The expanded definition is defined in the CEC Charter and includes hospitalizations that are determined to be heart failure hospitalizations by a CEC consensus yet may not have all of the documentation available.

5.2. M&M endpoints with follow-up ending at 6, 12 and 24 months

Over the course of follow-up in the trial, there were many unexpected changes that occurred in patient care that could confound the evaluation of the benefit of Barostim therapy. These include the approval and adoption of pharmaceutical and device-based treatments for heart failure (e.g., ARNI, SGLT2, Cardiac Contractility Modulation), as well as changes in care patterns and practices due to COVID-19. To better understand the impact of these event, the analyses on the morbidity and mortality endpoints will be repeated assuming that follow-up stops at 6 months, 12 months and 24 months. The same endpoint definitions as described above and analysis methods in Section 2.1.4 will be repeated while censoring follow-up time and events at 6, 12, and 24 months.

5.3. Alternative Statistical Methods for Analyzing Rates

To further incorporate the impact of the timing that the events occur over follow-up, alternative methods of analysis on the endpoints will be performed. These will include, at a minimum, the Anderson Gill method for recurrent events, and semi-parametric joint frailty models (JFM).

The JFM analysis is intended to aid the interpretation of the primary negative binomial model analysis by providing insight into the contribution of censoring events and recurrent heart failure hospitalization events to the overall (marginal) treatment effect estimated by the negative binomial model analysis. This JFM has a demonstrated ability to provide unbiased estimate of the treatment effect on recurrent heart-failure hospitalizations in the presence of dependent censoring (i.e., censoring being more likely in subjects at higher risk of heart-failure hospitalization). In the JFM, the dependence of censoring and recurrent events is accounted for by a frailty term, which measures a patient-specific tendency to have an increased likelihood of both types of events. The model assumes that a patient's censoring and recurrent events are independent of each other after conditioning on the frailty term. The JFM will be implemented using piecewise constant hazards for the recurrent events and censoring events. This method may be implemented by Gaussian quadrature using SAS PROC NLMIXED24. Piecewise constant hazard time intervals will be defined prior to performing analyses comparing treatment arms. For each hazard ratio comparing the two arms, a 95% confidence interval and p-value will be calculated using the Wald method. The JFM analysis will be repeated with only death as a dependent censoring event and transplant/LVAD as non-informative censoring events.

5.4. Reduction of the Components of the Endpoint

The objective is to demonstrate that treatment with the BAROSTIM NEO System, relative to medical management:

- Reduces the rate of hospitalizations for heart failure*
- Reduces the rate of hospitalizations for heart failure* or emergency room visits for heart failure
- Reduces cardiovascular mortality
- Reduces LVAD or heart transplant
- Reduces LVAD
- Reduces heart transplant
- Reduces cardiovascular mortality, LVAD or heart transplant

*May include hospitalization equivalent events, as well as heart failure events using the expanded definition.

For the analyses that include hospitalizations and hospitalization equivalent events only, the primary analysis will be a semi-parametric frailty model (as described in Section 4.2) that adjusts for the competing risk of the censoring events (e.g., cardiovascular mortality and/or LVAD and/or heart transplant).

For the analyses that include censoring events only, the analyses will be evaluated descriptively with Kaplan-Meier time-to-event methods. The event date will be the date of censoring event. Living subjects will be censored at the date of last contact. Subjects that die due to non-cardiovascular causes will be censored at the date of death. The estimate of survival and corresponding 95% log-log confidence interval as well as the number of subjects at risk and number of subjects that die will be presented at 6 months, 12 months, 24 months, and annually thereafter. The hazard ratio between arms and its 95% confidence interval will be estimated by a proportional hazard's regression model.

5.5. Win Ratio – M&M Only

A Win Ratio analysis[4] will be performed using the components of the morbidity and mortality endpoints with the following hierarchy.

1. CV Death
2. Heart transplant or LVAD
3. Number* of hospitalizations for Heart Failure

The Win Ratio analysis will use all pairs of subjects between the two study arms to determine the number of wins and losses for the treatment versus control arm.

For each pair, follow-up is defined by the shorter of the follow-up periods. For example, if one patient has 1 year of follow-up and the 2nd has 2 years, the follow-up time used for comparison is 1 year. If the first subject in a pair is censored prior to an event time

for the second subject in a pair, that pair is a tie. This produces the shared duration of follow-up for each pair, avoiding bias due to differential follow-up.[5]

The p-value comparing the two arms will be calculated using the Schoenfeld-Finkelstein method. A 95% confidence interval for the Win Ratio will be calculated using the bootstrap method.

5.6. Win Ratio Expanded – HF Therapy

A Win Ratio analysis[4] will be performed using the components of the morbidity and mortality endpoints as well as intensification of therapeutic interventions for heart failure. This analysis will follow the same methodology as the Win Ratio above. The following hierarchy will be used:

4. CV Death
5. Heart transplant or LVAD
6. Number of hospitalizations for Heart Failure
7. Intensification of therapeutic interventions for heart failure as defined in Table 3.

5.7. Reduction in Time Spent CV Death / HF Hospitalization

Reduction of the percent of follow-up time spent either dead from a cardiovascular cause and/or in the hospital for worsening heart failure between the two arms.

This analysis will be performed over all follow-up time, and at 6-, 12- and 24- months follow-up time.

5.8. Clinical Stability

The clinical stability endpoint categorizes subjects as “improved”, “Stayed the same” or “worsened” based on their mortality status, heart failure hospitalization status and NYHA class at 12 months. The number of HF hospitalization (HFH) in the 12 months prior to enrollment will be compared to the subjects’ post randomization HFHs through 12 months. A subject with less than 12 months follow-up will not be evaluable for this endpoint unless the subject died or experienced more HFH during follow-up as compared to the pre-enrollment 12 months (i.e., “worsened”). A subject will be classified as “worsened” if either the subject died prior to 12 months, the subject experienced more HFH through 12 months follow-up than the pre-enrollment 12 months, OR the subject has a higher NYHA class at 12 months as compared to baseline. A subject will be classified as “improved” if the subject did not worsen AND either had fewer HFH through 12 months follow-up as compared to pre-enrollment 12 months OR had an improvement in NYHA class at 12 months. A subject who neither worsened nor improved and who is evaluable on all 3 parameters through 12 months will be classified as “stayed the same”. The p-value comparing the two arms will be calculated using the proportional odds method assessing a shift in the distribution of the ordinal outcome. A

95% confidence interval for the proportional odds will be calculated using the bootstrap method.

5.9. Time to First Event

First of any component event (cardiovascular death, hospitalization for heart failure, cardiac assist device, and heart transplant) will be evaluated descriptively with Kaplan-Meier time-to-event methods. The event date will be the date of the first component event. Subjects without a component event will be censored at the date of last contact. The estimate of event free survival and corresponding 95% log-log confidence interval as well as the number of subjects at risk and number of subjects with an event will be presented at 6 months, 12 months, 24 months, and annually thereafter. The hazard ratio between arms and its 95% confidence interval will be estimated by a proportional hazard's regression model. Time to first event analyses will be repeated for each event type individually (separate analyses for cardiovascular death, hospitalization for heart failure, cardiac assist device, and heart transplant).

5.10. Reduction of All-Cause Mortality

All-cause mortality will be evaluated descriptively with Kaplan-Meier time-to-event methods. The event date will be the date of death. Living subjects will be censored at the date of last contact. The estimate of survival and corresponding 95% log-log confidence interval as well as the number of subjects at risk and number of subjects that die will be presented at 6 months, 12 months, 24 months, and annually thereafter. The hazard ratio between arms and its 95% confidence interval will be estimated by a proportional hazard's regression model.

5.11. Reduction of Heart Failure Mortality and Heart Failure Morbidity

The morbidity and mortality endpoint will be repeated with a revised definition for the terminal events that contribute to the event count outcome in the analysis. Only mortality events where the cause of death is heart failure will be counted as death endpoint events. All other deaths are censoring events, but not counted as primary endpoint events. The LVAD, heart transplant and heart failure morbidity events are treated the same as the main primary analysis (i.e., included in the analysis as primary events).

5.12. Reduction of HF Mortality

Mortality due to heart failure will be evaluated descriptively with Kaplan-Meier time-to-event methods. The event date will be the date of the heart failure death. Living subjects will be censored at the end of follow-up as defined in Section 2.1.1. Subjects that die due to non-heart failure causes will be censored at the date of death. The estimate of survival and corresponding 95% log-log confidence interval as well as the number of subjects at risk and number of subjects that die will be presented at 6 months, 12

months, 24 months, and annually thereafter. The hazard ratio between arms and its 95% confidence interval will be estimated by Cox proportional hazards regression.

5.13. Reduction in the Rate of within 30-day re-hospitalization for Heart Failure

Incidence of re-hospitalization for heart failure within 30-days by treatment arm will be analyzed descriptively.

5.14. Additional Mortality and Heart Failure Morbidity Definitions

Additional definitions of morbidity and mortality events will be explored as sensitivity analyses. These sensitivity analyses include pre-defined heart failure definitions that vary in severity and intensity of symptoms and/or heart failure treatments. These definitions may reflect recent changes and approaches to the care of heart failure patients.

6. ANCILLARY ENDPOINTS

The following endpoints will be evaluated for exploratory and supportive purposes. No formal hypotheses are provided.

6.1. New York Heart Functional Class (NYHA)

NYHA Class will be collected at baseline, 6, 12, 18 and 24 months. The frequency and percentage of subjects who improved in NYHA Class from baseline at each follow-up time point by treatment arm will be presented. A generalized estimating equation (GEE) repeated measures logistic model with a compound symmetry working correlation will be used to account for within subject correlation over time and to estimate the odds ratio for improved NYHA Class between the randomized arms over all times at which the endpoint is collected adjusted for baseline.

6.2. Minnesota Living With Heart Failure (MLWHF)

MLWHF questionnaire is collected at baseline and at 6, 12, 18 and 24 months. The mean improvement in MLWHF between baseline and at each follow-up time point by treatment arm will be presented with means and confidence intervals. A generalized estimating equation (GEE) repeated measures model with a compound symmetry working correlation will be used to account for within subject correlation over time and to estimate the difference between the randomized arms over all times at which the endpoint is collected adjusted for baseline.

6.3. EuroQol 5-Dimension (EQ-5D)

EQ-5D questionnaire is collected at baseline and at 6, 12, 18 and 24 months. Changes from baseline will be calculated at each follow-up time point for the index score and visual analogue score within each treatment arm. The mean improvement in EQ-5D between baseline and at each follow-up time point by treatment arm will be presented

with means and confidence intervals. A generalized estimating equation (GEE) repeated measures model with a compound symmetry working correlation will be used to account for within subject correlation over time and to estimate the difference between the randomized arms over all times at which the endpoint is collected adjusted for baseline.

6.4. Procedure Related Complications - Implant and 30 days Post-implant

Incidence of procedure related complications occurring between implant and 30 days post-implant will be analyzed descriptively with frequency counts and percentages for the safety population. The numerator will include the number of subjects with at least one procedure related complication occurring between implant and 30 days post-implant. The denominator will include subjects who have had an implant, or an implant attempted.

6.5. System Related Complications – 30 day to 6 months Post-implant

Incidence of system related complications occurring between 30 days and 6 months post-implant will be analyzed descriptively with frequency counts and percentages for the safety population. The numerator will include the number of subjects with at least one system related complication occurring between 30 days and 6 months post-implant. The denominator will include subjects who have had an implant, or an implant attempted.

7. COVID-19

On March 13, 2020, the US announced a national emergency declaration due to the COVID-19 pandemic[6]. Due to the changes in care patterns and accessibility to healthcare, as well as the interruption of logistics at trial sites, COVID-19 has had a significant impact on the execution, results and interpretation of clinical trials. These issues have been discussed globally by medical groups as well as governmental agencies.[7,8,9,10] In addition, there are examples from recent clinical trial that have demonstrated the potential impact of COVID-19 on the effectiveness of therapies.[11,12]

To adjust for, measure and understand the impact of COVID-19, several changes, including the collection of key data, were prospectively made to the trial in early 2020, including:

- Completion of randomizations at a total of 467 subjects randomized (13 less than then 480 randomized originally planned).
- To minimize the exposure of subjects, site personnel and CVRx personnel to COVID-19, protocol required follow-up visits could be completed via telephone. This reduced or delayed the completion of testing required at certain follow-up visits (e.g., six minute hall walks, blood draws). These deviations were specifically collected and noted as related to COVID-19.
- All site monitoring is done remotely until the site, CVRx and the monitor agree that an in-person meeting is safe and necessary.
- Device programming, if possible and appropriate, may be done remotely.
- A COVID-19 case report form (see Section 8 for copy of form), dedicated to understanding the impact of COVID-19 on subjects in the trial, was designed and implemented based on a template developed by the Heart Failure Collaboratory (www.hfcollaboratory.com). This form is required for all subjects for all follow-up visits starting on or after 01April2020, depending on IRB approval and/or site implementation. The form includes questions such as: “Since the last visit did the subject FEEL THE NEED to go to an emergency department or hospital for their heart failure BUT DECIDED NOT TO because of concerns about COVID-19”.
- The CEC Charter was updated to collect COVID-19 relatedness to endpoint. Specifically, for each adjudicated event the following evaluation of the impact of COVID was documented:
    - If COVID-19 is the primary reason for hospitalization, the committee indicates a secondary reason for hospitalization.
    - If COVID-19 is not the primary reason for hospitalization, the committee indicates whether it was a contributing factor.
    - If a determination cannot be made regarding the primary reason, the committee marks both as the primary reason.

Despite, or possibly because of, these changes, the impact of COVID-19 on the overall Barostim Therapy (BAT) treatment effect in this trial is unknown. Therefore, the following approach has been outlined to evaluate and measure any potential impact on the M&M endpoint. Sixty percent (60%) of the approximate 320 expected primary events occurred prior

to March 13, 2020. Thus, approximately 40% of the expected events will accrue during the COVID pandemic. This section outlines additional analyses of the M&M endpoints that may be conducted, as well as the qualitative summaries that will be provided in the assessment of the impact of the COVID-19 pandemic on the M&M data and conclusions.

To evaluate the impact of COVID-19 on the morbidity and mortality endpoints, an indicator variable reflecting events pre vs post the March 13, 2020 date will be included in the regression model and evaluated using a treatment-by-COVID-19 interaction. A p-value less than 0.15 for the interaction term will be considered evidence of a significant interaction between the treatment effect pre COVID-19 vs post COVID-19 (consistent with the interaction p-value for site and gender defined in the main Statistical Analysis Plan 360050-001). However, an insignificant interaction p-value ($p \geq 0.15$) does not necessarily mean that there is no significant difference in the magnitude of the benefit from Barostim therapy pre vs post COVID-19, primarily due to potentially low statistical power for evaluating the interaction term. Thus, the interaction p-value will be included as part of the totality of information used to characterize and evaluate the potential impact of COVID-19.

A stratified analysis, using the negative binomial model as specified in the primary analysis, may also be performed to further evaluate any difference in treatment effect pre-March 13, 2020, versus post-March 13, 2020. Separate estimates of the treatment effect will be made with data pre-March 13, 2020, and post-March 13, 2020. The follow-up time for the first model will be censored on March 12, 2020, while the follow-up start time for the second model will be March 13, 2020. The Anderson-Gill method for recurrent events model may also be used along with a time-varying covariate for follow-up pre- and post- March 13, 2020. The significance of the time-varying covariate as well as the interaction with treatment group would be reported. As with the negative binomial model, a p-value for the interaction term less than 0.15 will be considered evidence of a significant interaction between the treatment effect pre COVID-19 vs post COVID-19.

If there is evidence that the treatment effect is different pre- vs post-March 13, 2020, then the benefit observed pre-March 13, 2020 would be more consistent with the assumptions and definitions of the study endpoints. This includes the definition and assumed event rate for heart failure hospitalization, as it has been observed that patients have shown a reluctance to interact with the health care setting during the COVID-19 pandemic, some by reducing their risk of hospitalization by changing their behaviors and others by receiving care remotely to replace what may have previously been a hospitalization (e.g. acute medication changes).[11]

In addition to the analyses specified above, other additional relevant information will be considered, such as:

- Number and percent of pre- vs post- March 2020 information (events, follow-up time, rate)
- Follow-up visit compliance pre- vs post- March 2020
- Comparison of other definitions of heart failure on event rates between treatment arms pre- vs post- March 2020, such as
    - Expanded definition of heart failure (see CEC Charter)
    - Unscheduled clinic visits

  - Acute changes in medications
- Evaluate the impact using the responses on the COVID-19 case report form. This includes summarizing, or incorporating in the analyses, the instances where subjects reported the need to go to an emergency department or hospital for their heart failure but decided not to because of concerns about COVID-19.
- Evaluate the information regarding the CEC determined impact of COVID-19 on events, such as HF hospitalizations where CEC determined the event was a study endpoint, but COVID-19 was considered a "Co-primary" or a contributing factor, or where COVID-19 was the primary reason for hospitalization, but heart failure was the secondary reason.
- Evaluate all COVID related adverse events and deaths
- Evaluate COVID related deviation and the impact on the trial and results

There are three interim analyses schedule at approximately 140 (44%, Jan 2020), 180 (~56%, Oct 2020), 240 (~75%, Oct 2021) primary endpoint events, and a final at ~320 events (100%, TBD). Thus, the first and possibly much of the second interim analyses primarily include data that were not impacted by COVID-19. As such, these analyses may also provide insight into the impact of COVID-19 and will be provided in the PMA Supplement submission.

The totality of evidence available will be considered when evaluating the impact of COVID. This includes all evidence of a differential treatment effect of Barostim therapy pre- vs. post-March 13. This includes, at a minimum, the analyses specified above as well as the qualitative assessments of the impact of the COVID-19 pandemic on the study based on the relevant information collected above.

8. ASSUMPTIONS FOR THE PRIMARY MORBIDITY AND MORTALITY ENDPOINT

8.1. Power and Type I Error Calculation

Since a subgroup of patients with partial data has been previously analyzed, the relevant power and type I error for the future interim analyses is the conditional power and type I error, accounting for the fact that partial data has been previously analyzed. That previous analysis did not incur any alpha spending since there were no criteria for stopping the trial for success based on that analysis.

The conditional power and type I error are calculated by Monte Carlo methods using simulations of hypothetical trial data simulated according to the joint negative binomial model with parameters estimated based on the trial data through 03October2018.

Prospective follow-up and events from 03OCT2018 through trial completion at 320 events are simulated using a negative binomial model. Multiple hypothetical sets of future events are simulated and the pooled data, including observed and simulated data, is analyzed using the joint negative binomial model.

Pooling the data is equivalent to 100% borrowing. Pooling is performed in order to maintain the interpretability of the analysis, due to the intractability of separating the cumulative follow-up data into separate time periods while maintaining interpretability. The impact of 100% borrowing on type 1 error is accounted for by using a sufficiently rigorous nominal alpha such that conditional type 1 error is well controlled at less than 5%.

Programs will be provided to FDA that implement the Monte Carlo calculations.

Using 20,000 simulations for the null and 1000 simulations for the alternative under the assumptions below, the conditional power is 86.0% and the type I error is 4.4% when using an alpha spending function with a cumulative alpha level of 0.01. The stricter alpha level is sufficient to offset the tendency of the interim data to inflate the type 1 error. Using a group sequential spending function with a cumulative alpha of 0.01 results in controlling the type 1 error to less than 5% across a range of scenarios.

The base case uses the following assumptions for the conditional power and type I error calculations:

1. Control arm event rate of 0.4 events per patient year
2. Power calculated assuming an expected relative reduction in the endpoint event rate of $1-\exp(\beta 1) = 0.35$
3. Type I error calculated assuming an expected relative reduction in the endpoint event rate of $1-\exp(\beta 1) = 0$, equivalent to no treatment effect
4. Statistical significance threshold given by the Hwang-Shih-DeCani spending function with $\gamma = -3$

5. Sample size of 331 subjects with baseline NT-proBNP < 1600, consisting of 259 subgroup subjects randomized through 03October2018 and an additional 72 subjects to be prospectively randomized.
6. Parameter describing the relationship between events in the year prior to randomization and the expected rate of events during follow-up (beta) of 1.8
7. Parameter describing the probability that an event is a censoring event (p) of 0.18
8. Parameter describing the dispersion of the negative binomial data (alpha) of 1.7

8.2. Operating Characteristics

The operating characteristics under the group-sequential evaluation of the morbidity and mortality results are summarized below. In addition to varying the treatment effect assumption under the assumptions of the base case, the power and type 1 error of the trial were also evaluated under a variety of scenarios where other parameters differ from the base case assumptions.

Table 4 shows the variety of scenarios explored under the null effect size to demonstrate that the type 1 error is controlled to less than 5% even if accrual, event rate, censoring probability, and dispersion differ from the data observed as of 03October2018. Compared to the scenarios explored in the original trial design, the variation in parameters is lower due to the previous information observed. The value of beta is not varied since a very high proportion of patients have known data regarding the events in the year prior to enrollment.

Table 4: Null Scenarios

	Description	λ	θ	p	α	Accrual
A1	Base	0.4	0	0.18	1.7	100
A2	Slow accrual	0.4	0	0.18	1.7	**75**
A3	Fast accrual	0.4	0	0.18	1.7	**125**
A4	Low event rate	**0.35**	0	0.18	1.7	100
A5	High event rate	**0.45**	0	0.18	1.7	100
A6	Low terminal fraction	0.4	0	**0.1**	1.7	100
A7	High terminal fraction	0.4	0	**0.3**	1.7	100
A8	Large overdispersion	0.4	0	0.18	**1.4**	100

Note: Null scenarios in terms of the control arm event rate λ, the device effect θ (always zero in the null scenario), the fraction of events p that are terminal, the negative binomial overdispersion parameter α, and accrual rate of the additional 72 randomized subjects.

Table 5 shows the estimated type 1 error rates across these scenarios based on 20,000 simulations per scenario. None of the estimated type 1 errors exceeds 5%.

Table 5: Conditional Type 1 Error Estimates across Scenarios with Nominal Alpha of 0.01

	Description	MC estimate	MC s.e.	MC 95% CI	
A1	Base	0.0436	0.0014	0.0408	0.0465
A2	Slow accrual	0.0422	0.0014	0.0394	0.0450
A3	Fast accrual	0.0448	0.0015	0.0420	0.0477
A4	Low event rate	0.0478	0.0015	0.0448	0.0507
A5	High event rate	0.0407	0.0014	0.0380	0.0435
A6	Low terminal fraction	0.0440	0.0015	0.0412	0.0468
A7	High terminal fraction	0.0455	0.0015	0.0426	0.0484
A8	Large overdispersion	0.0452	0.0015	0.0423	0.0480

Table 6 shows the alternative scenarios that were explored, varying the assumptions about relative rate between randomized arms (effect size), accrual rate, event rate, and dispersion to match the scenarios under the null.

Table 6: Alternative Scenarios

	Description	λ	θ	p	α	Accrual
B1	Null	0.4	**0**	0.18	1.7	100
B2	25% effect size	0.4	**log(0.70)**	0.18	1.7	100
B3	35% effect size	0.4	**log(0.65)**	0.18	1.7	100
B4	45% effect size	0.4	**log(0.55)**	0.18	1.7	100
B5	Slow accrual	0.4	log(0.65)	0.18	1.7	**75**
B6	Fast accrual	0.4	log(0.65)	0.18	1.7	**125**
B7	Low event rate	**0.35**	log(0.65)	0.18	1.7	100
B8	High event rate	**0.45**	log(0.65)	0.18	1.7	100
B9	Low terminal fraction	0.4	log(0.65)	**0.1**	1.7	100
B10	High terminal fraction	0.4	log(0.65)	**0.3**	1.7	100
B11	Large overdispersion	0.4	log(0.65)	0.18	**1.4**	100

Note: Alternative scenarios in terms of the control arm event rate λ, the device effect θ (always zero in the null scenario), the fraction of events p that are terminal, the negative binomial overdispersion parameter, and accrual rate.

Since the information fraction is estimated based on the observed event rate and dispersion, the boundaries are characterized by the mean alpha spent at each interim analysis. The z-boundaries and nominal p-values needed for statistical significance at each interim analysis are also displayed in Table 7.

Table 7: Group Sequential Parameters by Interim Analysis (Mean across Simulations under B1)

	Total	@140 events	@180 events	@240 events	@320 events
Alpha Spent	0.01	0.0019	0.0011	0.0024	0.0046
Z-Boundary	NA	2.9032	2.8781	2.6713	2.4335
Nominal p-value	NA	0.0018	0.002	0.0038	0.0075

Finally, the conditional power for the trial is shown in Table 8. The table shows the probability of reaching statistical significance at each interim analysis, and cumulatively across all interim analyses, across the scenarios described in Table 6.

Table 8: Conditional Power Error Estimates across Scenarios with Nominal Alpha of 0.01

	Description	Total	@140 events	@180 events	@240 events	@320 events
B1	Null	0.0436	0.0098	0.0063	0.0096	0.018
B2	25% effect size	0.553	0.118	0.089	0.150	0.196
B3	35% effect size	0.860	0.275	0.163	0.236	0.186
B4	45% effect size	0.985	0.560	0.208	0.162	0.055
B5	Slow accrual	0.855	0.271	0.138	0.253	0.193
B6	Fast accrual	0.866	0.306	0.169	0.205	0.186
B7	Low event rate	0.871	0.285	0.198	0.226	0.162
B8	High event rate	0.855	0.283	0.155	0.219	0.198
B9	Low terminal fraction	0.849	0.270	0.167	0.242	0.170
B10	High terminal fraction	0.838	0.296	0.174	0.209	0.159
B11	Large overdispersion	0.839	0.308	0.169	0.195	0.167

8.3. Summary

Assuming a 35% relative reduction between the arms, the final sample size for the M&M endpoint is 331 randomized subjects with NT-proBNP<1600 pg/ml, resulting in a condition power of 86% and a type 1 error of 4.4%. The 331 subjects are a combination of 259 previously randomized and 72 prospectively randomized subjects. Events accumulated up to the 03OCT2018 interim analysis combined with all future events will be used to evaluate the relative reduction between the treatment arms based on a negative binomial model. The pooling is to maintain the interpretability of the data. The impact of the pooling was accounted for by using a sufficiently rigorous nominal alpha such that conditional type 1 error is well controlled at less than 5% over a variety of null scenarios.

Interim analyses are planned at 140, 180, 240, and 320 primary endpoint events, with a final analysis after 320 events are accumulated (as needed). A group sequential spending function with a strict cumulative alpha level of 0.01 was used to offset the tendency of the interim data to inflate the type 1 error.

9. COVID-SPECIFIC CASE REPORT FORM

SECTION A. ADMINISTRATIVE SECTION

Subject Initials: ___ ___ ___ Visit Interval: _____

SECTION B. COVID-LIKE ILLNESS SYMPTOMS

Since the last visit has the subject experienced any cold, flu or COVID-19 symptoms?

☐ No (skip to section C)
☐ Yes, date of onset: _____/_____
MMM/YYYY

If yes, check all symptoms that apply:

☐ Fevers or chills
☐ New or Worsening Cough
☐ Productive ☐ Dry
If yes to cough, indicate severity
☐ Constant ☐ Occasional, several per hour
☐ New or worsened shortness of breath
☐ Diarrhea
☐ Altered or reduced sense of smell or taste
☐ Muscle aches/Severe Fatigue
☐ Chest pain or tightness
☐ Sore throat
☐ Nausea or vomiting

SECTION C. COVID-LIKE ILLNESS TESTING

Since the last visit has the subject been tested for COVID-19?

☐ No
☐ Yes, date of test: _____/_____
MMM/YYYY

If yes, test results:
☐ Positive ☐ Negative ☐ Unknown

Since the last visit has the subject been tested for COVID-19 antibodies?

☐ No
☐ Yes, date of test: _____/_____
MMM/YYYY

If yes, test results:
☐ Positive ☐ Negative ☐ Unknown

Since the last visit has the subject been vaccinated for COVID-19?

☐ No
☐ Yes, date of vaccine: _____/_____
MMM/YYYY

Since the last visit has the subject been tested for Influenza ("flu")?

☐ No
☐ Yes, date of test: _____/_____
MMM/YYYY

If yes, test results:
☐ Positive ☐ Negative ☐ Unknown

Has the subject been vaccinated for Influenza ("flu")?	☐ No ☐ Yes, date of vaccine: _____/_____ MMM/YYYY
SECTION D. COVID-LIKE ILLNESS EXPOSURE	
Since the last visit has the subject been told that they might have had COVID-19/have symptoms suggestive of COVID-19??	☐ No ☐ Yes ☐ Unknown
Since the last visit has the subject been exposed to anyone with known or suspected COVID-19?	☐ No ☐ Yes ☐ Unknown
Since the last visit has the subject been told that they might have had the "flu" or influenza?	☐ No ☐ Yes ☐ Unknown
SECTION E. EFFECTS OF COVID PANDEMIC ON PATIENT'S HEALTHCARE INTERACTIONS	
Since the last study visit, did the subject **feel the need** to go to an emergency department or hospital for their heart failure **but decided not to** because of concerns about COVID-19? If yes, how did they seek care (check all that apply): ☐ Telemedicine visit ☐ In-person ☐ Clinic ☐ Urgent Care ☐ Subject did not change hospital/ER use due to COVID-19 ☐ Other, specify: _____	☐ No ☐ Yes ☐ Unknown
Since the last study visit, did the subject have a cardiology/HF related appointment cancelled/rescheduled due to COVID-19 pandemic?	☐ No ☐ Yes, how many? _____ ☐ Unknown
Since the last study visit, did the subject have any cardiology/HF related telemedicine visit due to COVID-19 pandemic?	☐ No ☐ Yes, how many? _____ ☐ Unknown
Since the last study visit, did the subject have a cardiology/HF procedure cancelled/rescheduled due to COVID-19 pandemic?	☐ No ☐ Yes, how many? _____ ☐ Unknown
SECTION F. EFFECTS OF COVID PANDEMIC ON PATIENT'S MEDICATIONS	
Since the last visit, did any of their heart failure medications change or stop, even for a short time? If yes, enter change into medication eCRF If yes, why: ☐ Instructed by Doctor ☐ Self-discontinued ☐ Unknown ☐ Other: _____	☐ No ☐ Yes ☐ Unknown

Since the last visit, was the subject prescribed any medications for COVID-19? If yes, what medications (generic name):	☐ No ☐ Yes ☐ Unknown
SECTION G. EFFECTS OF COVID PANDEMIC ON PATIENT'S LIFESTYLE	
How has the subject's activity/exercise level changed due to COVID-19 pandemic?	☐ No change ☐ More activity/exercise ☐ Less activity/exercise
How has the subject's smoking habits changed due to COVID-19?	☐ Does not smoke ☐ No change ☐ Smokes more ☐ Smokes less
How has the subject's alcohol drinking habits changed due to COVID-19?	☐ Does not drink ☐ No change ☐ Drinks more ☐ Drinks less
How has the subject's eating habits changed due to COVID-19?	☐ No change ☐ More healthy ☐ Less healthy

[1] Mütze T, Glimm E, Schmidli H, Friede T (2017). Group sequential designs for negative binomial outcomes. *Statistical Methods in Medical Research,* DOI 10.1177/0962280218780538

[2] BeAT-HF SAP Rev E, Appendix D

[3] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.

[4] Pococks, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal*, 2012; 33 (2) 176-182; DOI: 10.1093/eurheartj/ehr352

[5] Redfors et al, The win ratio approach for composite endpoints: practical guidance based on previous experience, EHJ, 2020 41, 4391-4399.

[6] Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak. March 31, 2020. https://trumpwhitehouse.archives.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/

[7] Anker SD, Butler J, Khan MS, et al. Conducting clinical trials in heart failure during (and after) the COVID-19 pandemic: an Expert Consensus Position Paper from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J 2020; 41: 2109–17.

[8] European Medicines Agency. Guidance on the management of clinical trials during the COVID-19 (coronavirus) pandemic. Version 4. April 2, 2021. https://ec.europa.eu/health/sites/default/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf.

[9] US Food and Drug Administration. Conduct of clinical trials of medical products during the COVID-19 public health emergency. March 2020. https://www.fda.gov/media/136238/download.

[10] Psotka MA, Abraham WT, Fiuzat M, et al. Conduct of clinical trials in the era of COVID-19: JACC Scientific Expert Panel. J Am Coll Cardiol 2020; 76: 2368–78.

[11] Lindenfeld J, Abraham WT, Maisel A, Zile M, Smart F, Costanzo MR, Mehra MR, Ducharme A, Sears SF, Desai AS, Paul S, Sood P, Johnson N, Ginn G, Adamson PB. Hemodynamic-GUIDEd management of Heart Failure (GUIDE-HF). Am Heart J. 2019 Aug;214:18-27. doi: 10.1016/j.ahj.2019.04.014. Epub 2019 May 3. PMID: 31150790.

[12] Ponikowski P, et. al. Ferric carboxymaltose for iron deficiency at discharge after acute heart failure: a multicentre, double-blind, randomised, controlled trial. Lancet. 2020 Dec 12;396(10266):1895-1904. doi: 10.1016/S0140-6736(20)32339-4. Epub 2020 Nov 13. Erratum in: Lancet. 2021 Nov 27;398(10315):1964. PMID: 33197395.