



# **Statistical Analysis Plan**

## **SHINE**

### **Multi-electrode Radiofrequency Balloon Catheter use for the Isolation of the Pulmonary Veins**

**Protocol# BWI\_2017\_01 (v 5.0)**

Version 1.0

Oct 02, 2018

Sponsor:                      BIOSENSE WEBSTER, INC.  
33 Technology Dr.,  
Irvine, CA 92618  
USA

Biosense Webster, Inc

CONFIDENTIAL

## Table of Contents

1	Study Design .....	7
2	Treatment Assignment .....	7
3	Randomization and Blinding Procedures .....	7
4	Measures to Minimize bias .....	7
5	Interval Windows .....	8
6	Study Endpoints .....	8
6.1	Primary Endpoints and Associated Hypotheses .....	8
6.1.1	Primary Safety Endpoints .....	8
6.1.2	Primary Effectiveness Endpoints .....	8
6.2	Secondary Endpoints .....	9
6.2.1	Secondary Safety Endpoints .....	9
6.2.2	Secondary Effectiveness Endpoints .....	9
6.3	Additional Endpoints .....	10
6.4	Health Economic Data .....	10
7	Level of Significance .....	10
8	Analysis Sets .....	11
8.1	Main Study .....	11
8.2	Roll-in .....	11
9	Sample Size Justification .....	11
10	Data Monitoring Committee .....	12
11	Statistical Analysis Methods .....	13
11.1	General Conventions .....	13
11.2	Subject Disposition .....	13
11.3	Demography and Baseline Characteristics .....	13
11.4	Analysis of Primary Endpoints .....	13
11.4.1	Primary Effectiveness Endpoint .....	13
11.4.2	Primary Safety Endpoint .....	13
11.4.3	Interim Analysis .....	14
11.4.4	Handling of Missing Data .....	14
11.4.5	Sensitivity Analyses .....	15
11.4.6	Subgroup Analyses .....	16
11.5	Analysis of Secondary Endpoints .....	16

Biosense Webster, Inc

11.5.1	Secondary Safety Analysis .....	16
11.5.2	Secondary Effectiveness Analysis .....	16
11.5.2.1	Secondary Effectiveness Endpoints of Effectiveness Success .....	16
11.5.2.2	PVI and Non-PV Triggers Touch-Up with Focal Catheter .....	17
11.5.3	NAE Endpoints .....	17
11.6	Analysis of Procedural Data .....	18
11.7	Analysis of Atrial Fibrillation Effect on QualiTy-of-life Questionnaire .....	18
11.8	Analysis of Health Economic Data .....	19
12	Appendix – Adaptive Design Simulation Report .....	20
13	References .....	27

**Multi-electrode Radiofrequency Balloon Catheter use for the Isolation of the Pulmonary Veins (SHINE)**

**Statistical Analysis Plan Version: 1.0**

**Corresponding to: Protocol Version: 5.0**

The following individuals have reviewed this version of the Statistical Analysis Plan and are in agreement with the content:

**Signature Page**

[Redacted Signature]

[Redacted Signature]

[Redacted Signature]

[Redacted Signature]

[Redacted Signature]

[Redacted Signature]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

**List of Abbreviations**

AAD	antiarrhythmic drug
AF	atrial fibrillation
AFEQT	Atrial Fibrillation Effect on QualiTy-of-life
AFL	atrial flutter
AT	atrial tachycardia
DMC	Data Monitoring Committee
mITT	Modified Intent-To-Treat
MoCA	Montreal Cognitive Assessment
NAE	neurological assessment evaluable subgroup
NIHSS	National Institutes of Health Stroke Scale
PAE	primary adverse event
PAF	paroxysmal atrial fibrillation
PP	per-protocol
PV	pulmonary vein
PVI	pulmonary vein isolation
RF	radiofrequency
SADE	Serious Adverse Device Effects
SAE	serious adverse event

## 1 STUDY DESIGN

This clinical investigation is a prospective, multicenter, single arm clinical evaluation utilizing the Biosense Webster Multi-Electrode Radiofrequency Balloon catheter and the Biosense Webster Multi-Electrode Circular Diagnostic catheter.

An adaptive Bayesian design <sup>[1]</sup> will be used to determine the sample size. A maximum of 230 evaluable subjects with symptomatic paroxysmal atrial fibrillation (PAF) may be enrolled in the main study phase. Subjects enrolled in the main study phase will be evaluated at 7 days, 1, 3, 6 and 12 months following the index procedures.

To minimize the learning curve effect on the evaluation of safety and effectiveness of the multi-electrode RF balloon catheter, the first 1 to 3 roll-in subjects will be assigned to each ablating physician in this study. Based on a maximum number of 20 sites, and assuming an average of 2 ablating physicians per site, up to an estimated 60 roll-in subjects will be included in the study.

A focused neuropathological evaluation will be integrated within the Main Study. A subset of subjects enrolled in the main study phase (40 subjects) will be included in the Neurological Assessment Evaluable (NAE) subgroup. Enrollment of NAE subjects may be terminated prior to achieving the target 40 subjects if study enrollment ends early after a planned interim look. NAE subjects will be assessed for incidences of symptomatic and asymptomatic pre-and post-ablation cerebral emboli with either an absence of CNS deficits (asymptomatic) or with emboli-associated neurological symptoms (symptomatic).

## 2 TREATMENT ASSIGNMENT

This is a single-arm study. The only treatment assigned is the Multi-Electrode Radiofrequency Balloon catheter.

## 3 RANDOMIZATION AND BLINDING PROCEEDURES

This study is a non-randomized single-arm study. Therefore, masking of treatment assignment for operators and subjects will not be performed.

## 4 MEASURES TO MINIMIZE BIAS

This study will employ several measures to minimize operational bias.

- Timing of the interim analyses for sample size selection will not be revealed to sites
- An independent statistician will be responsible for performing interim analyses
- Results from the interim analyses will not be shared with the Sponsor or sites unless the interim analysis results in a decision to stop enrollment

Biosense Webster, Inc

- Sponsor personnel directly involved in the conduct of the study will not have access to intermediate aggregated summaries of primary or secondary safety and effectiveness endpoint data until decision of study enrollment adaptation has been made

## 5 INTERVAL WINDOWS

Refer to protocol Table 8.7A.

## 6 STUDY ENDPOINTS

### 6.1 Primary Endpoints and Associated Hypotheses

#### 6.1.1 Primary Safety Endpoints

The primary safety endpoint is the incidence of early onset Primary Adverse Events (PAEs) (within seven (7) days of the initial mapping and ablation procedure). PAEs include the following AEs.

Death*	Stroke/CVA
Atrio-Esophageal Fistula*	TIA
Myocardial Infarction	Phrenic Nerve Paralysis
Cardiac Tamponade/perforation	Pulmonary Vein Stenosis*
Thromboembolism	Major Vascular Access Complication/Bleeding

\* Device or procedure related death, pulmonary vein stenosis and atrio-esophageal fistula that occur greater than one week (7 days) and less than or equal to 90 days post procedure are considered and analyzed as primary AEs.

The PAE rate will be compared against the performance goal of 15% by testing the following hypotheses:

$$H_0: p_S \geq 0.15 \quad \text{vs.} \quad H_A: p_S < 0.15,$$

where  $p_S$  is the PAE rate.

#### 6.1.2 Primary Effectiveness Endpoints

The primary effectiveness endpoint of this study is the proportion of subjects with acute procedural success defined as confirmation of entrance block in treated pulmonary veins (PV) after adenosine and/or isoproterenol challenge (with or without the use of a focal catheter). The hypotheses to be tested for this evaluation are:

$$H_0: p_E \leq 0.80 \quad \text{vs.} \quad H_A: p_E > 0.80,$$

where  $p_E$  is the rate of acute procedural success.



## 6.2 Secondary Endpoints

### 6.2.1 Secondary Safety Endpoints

- Incidence of individual PAE from the primary composite
- Incidence of Serious Adverse Device Effects (SADEs)
- Incidence of Serious Adverse Events (SAEs) within 7 days (early onset), 8-30 days (peri-procedural) and >30 days (late onset) of initial ablation procedure
- Incidence of non-serious adverse events
- Incidence of pre- and post-ablation asymptomatic and symptomatic cerebral emboli as determined by MRI evaluations
- Frequency, anatomic location, and size (diameter and volume) of cerebral emboli by MRI evaluations at baseline, post-ablation and during follow-up
- Incidence of new or worsening neurologic deficits post ablation and during follow up, compared to baseline.
- Summary of National Institutes of Health Stroke Scale (NIHSS) scores at baseline, post ablation and during follow-up.
- Summary of Montreal Cognitive Assessment (MoCA) scores at baseline, one-month follow-up and during further follow-up.
- Hospitalization for cardiovascular events (with hospitalization defined as prolonged stay  $\geq 2$  nights post index procedure or in-patient stay not concurrent with index procedure  $\geq 1$  calendar day)

### 6.2.2 Secondary Effectiveness Endpoints

- Percentage (%) of pulmonary vein isolation (PVI) touch-up by focal catheter among all targeted veins and by subject.
- Percentage (%) of subjects with use of focal catheter ablation for non-PV triggers
- 6-Month Documented Symptomatic AF Recurrence: Percentage (%) of subjects with freedom from documented, symptomatic atrial fibrillation (AF), atrial tachycardia (AT), or atypical (left side) atrial flutter (AFL) episodes (episodes >30 seconds on arrhythmia monitoring device from Day 91 to 180)
- 6-Month Documented AF Recurrence: Percentage (%) of subjects with freedom from documented, atrial fibrillation (AF), atrial tachycardia (AT), or atypical (left side) atrial

flutter (AFL) episodes (episodes >30 seconds on arrhythmia monitoring device from Day 91 to 180)

- 12-Month Documented Symptomatic AF Recurrence: Percentage (%) of subjects with freedom from documented, symptomatic atrial fibrillation (AF), atrial tachycardia (AT), or atypical (left side) atrial flutter (AFL) episodes (episodes >30 seconds on arrhythmia monitoring device from Day 91 to 365)
- 12-Month Documented AF Recurrence: Percentage (%) of subjects with freedom from documented, atrial fibrillation (AF), atrial tachycardia (AT), or atypical (left side) atrial flutter (AFL) episodes (episodes >30 seconds on arrhythmia monitoring device from Day 91 to 365)

### **6.3 Additional Endpoints**

- Total procedure time, ablation time, RF application time, balloon dwell time, time to effect (PVI)
- Number and time of RF applications per PV location
- Fluoroscopy time and dose

### **6.4 Health Economic Data**

- Index procedural workflow costs
- Hospital costs
- Quality of Life: Atrial Fibrillation Effect on QualiTy-of-life (AFEQT) questionnaire

## **7 LEVEL OF SIGNIFICANCE**

Each of the primary safety and effectiveness hypothesis tests will be performed at a one-sided 5.0% significance level. The safety of this device was initially assessed in a feasibility study, and a safety event rate of 2.6% (1 in 39 patients) was observed. Given the available data on subjects from the feasibility study it was concluded that the proposed study could provide sufficient evidence for safety of the device using a type-I error of 5%, rather than requiring the conventional two-sided 5%. Since study success is defined as meeting both endpoints and the type-I error for each test is controlled at 5.0%, the type-I error for claiming study success is controlled at 5.0%. In addition, we have demonstrated that the overall type-I error for study success is controlled at 5.0% for all hypothetical scenarios simulated including the ones where one or both safety and effectiveness rates are on the decision boundary.

## 8 ANALYSIS SETS

### 8.1 Main Study

For the analysis of study endpoints, the analysis populations defined in the following will be used:

- **Modified Intent-To-Treat (mITT) Population:** The mITT population will consist of enrolled subjects who meet eligibility criteria and have undergone insertion of the study catheters.
- **Safety Population (SP):** The SP will consist of all enrolled subjects who have undergone insertion of the study catheters.
- **Per Protocol (PP) Population:** The PP population is a subset of the mITT population and will include subjects who comply with the following criteria:
  - are enrolled and meet all eligibility criteria
  - have undergone RF ablation with study catheters
  - are treated for the study-related arrhythmia

**Neurological Assessment Evaluable (NAE) Population:** The NAE Population will include at least 40 subjects who are subset of the Per Protocol Population, consent and are eligible for the required neurological assessments, and have post-ablation MRI. Assessment of incidence of new lesions requires availability of pre-and post MRI. The subjects without pre-MRI but no lesion on post-MRI will be included in the NAE population. Assessment of incidence of new deficits requires availability of pre-and post-neurological evaluation. Enrollment in the NAE population may be terminated prior to achieving the target number of subjects if study enrollment ends early after a planned interim look.

### 8.2 Roll-in

- **Roll-In Population:** The Roll-In population will include all subjects who are enrolled in the roll-in phase and have the study catheter inserted and RF delivered. All endpoints will be analyzed for the roll-in population separately.

## 9 SAMPLE SIZE JUSTIFICATION

A Bayesian adaptive design will be utilized to select the final sample size of the trial. The sample size for the study is mainly driven by the primary safety endpoint. Sample size selection interim analyses will be performed when the mITT Population reaches 80, 130, and 180 subjects. Safety outcomes at 30 days will be used as a proxy for the primary safety endpoint at each

Biosense Webster, Inc

interim for determining the sample size. The predictive probabilities for the primary safety endpoint success will be used to determine whether the sample size at the time of the interim analysis will be sufficient or if the trial will continue to the full sample of 230. The final stopping sample size of 230 was determined by fixed design power calculations for the primary safety endpoint.

Trial simulations were performed to estimate the Type-I error rate and the power for the success of the safety endpoint under a range of assumptions for the true safety rate and the assumed enrollment schedule. It is assumed that the attrition rate for the study will be negligible. For accrual rates of 1.7, 3.4, 5.1, 8.5, 11.9, 17, 20.4, 23.8, 25.5, 25.5, 25.5, 25.5, 25.5, and 25.5 patients per month for months 1 to 15 respectively, and true safety rates of 0.0% to 9.0%, the power for showing success for the primary safety endpoint is estimated to be greater than or equal to 80.0%. For a safety rate of 9.5% this power decreases to approximately 73.0%. Based on the feasibility study results referenced in the study protocol, it is unlikely that the safety event rate is higher than 9.0%. The posterior probability of the safety rate being less than or equal to 9.0% is estimated to be 0.886. Therefore, the study is expected to be sufficiently powered for the primary safety endpoint. With an assumed true effectiveness rate of 96.0% and a performance goal of 80.0%, a minimum of 80 subjects will provide higher than 99.0% power for meeting the primary effectiveness endpoint. As a result, the power estimates for the primary safety endpoint closely estimate the overall power of the study.

Since the primary effectiveness analysis is performed at a 5.0% level Type-I error and the sample size is selected independent of the effectiveness endpoint, the Type-I error for the effectiveness hypothesis is controlled at 5.0%. To estimate the Type-I error for the primary safety endpoint, we considered the hypothetical scenario that the primary safety rate is on the decision boundary (i.e. equal to 15%). Based on 25,000 trial simulations, the estimated Type-I error for this scenario was 4.7%. As the primary safety rate moves away from the decision boundary under the null hypothesis, the Type-I error for the test decreases. Since the study success is defined as meeting both primary safety and effectiveness endpoints and because the Type-I error for each of these tests is controlled at 5.0%, the overall study Type-I error will be controlled at 5.0%. (Please refer to the Appendix for more details).

Based on the simulation results, the study is adequately powered to meet the primary safety and effectiveness endpoints and that the Type-I error for the overall trial success is controlled at 5%.

## **10 DATA MONITORING COMMITTEE**

A Data Monitoring Committee (DMC) will be constituted to monitor subject safety and provide guidance on study adaptation. The DMC charter will document roles and responsibilities of the committee and the independent statistician.

Biosense Webster, Inc



## **11 STATISTICAL ANALYSIS METHODS**

### **11.1 General Conventions**

Standard descriptive summaries for continuous data will include the number of observations with data, mean, standard deviation, median, 25% percentile, 75% percentile, minimum, and maximum values. For categorical data, count and percentages of observations will be provided. Summaries will be based on subjects without missing data.

### **11.2 Subject Disposition**

Disposition and accountability of the study subjects will be summarized descriptively for the subject categories defined in section 4.4 of the protocol.

### **11.3 Demography and Baseline Characteristics**

Subject demographics, medical history, AAD medication history and other baseline data will be summarized descriptively for all enrolled subjects as well as the safety, mITT, per protocol and NAE populations. Descriptive statistics will also be presented for roll-in subjects.

### **11.4 Analysis of Primary Endpoints**

#### **11.4.1 Primary Effectiveness Endpoint**

The acute procedural success rate will be compared to the performance goal of 80% by using the exact test for a binomial proportion at a one-sided significance level of 5%. If the one-sided 95% lower confidence bound for the primary effectiveness rate is greater than the performance goal of 80%, the study will be considered to have demonstrated effectiveness.

The per-protocol (PP) population will be used as the analysis population for the primary effectiveness endpoint. Subjects with missing effectiveness endpoints data will be excluded in the primary analysis. The primary effectiveness endpoint will also be analyzed in the roll-in subjects.

#### **11.4.2 Primary Safety Endpoint**

The PAE rate will be compared to a performance goal of 15% by using the exact test for a binomial proportion at a one-sided significance level of 5%. If the one-sided 95% upper confidence bound for the primary safety rate is less than the performance goal of 15%, the study will be considered to have demonstrated safety.

The Modified Intent-To-Treat (mITT) population will be used as the analysis population for the primary safety endpoint. Subjects without full follow-up required per protocol for the safety

endpoint assessment will be excluded in the primary analysis unless they have experienced a PAE. The primary safety endpoint will also be analyzed in the roll-in subjects.

#### **11.4.3 Interim Analysis**

Sample size selection interim analyses will be performed when the mITT population reaches 80, 130, and 180 subjects. Safety outcome at 30 days will be used as a proxy for the primary safety endpoint at each interim.

At the time of each interim analysis, predictive probabilities of success for safety will be calculated once using the current sample size and another time using the maximum sample size allowed. A uniform prior distribution will be assumed for the safety rate at the first interim analysis, and the distribution will be updated with the observed data at each subsequent interim analysis. The number of failures for the primary safety endpoint in subjects with incomplete information will be assumed to follow a beta-binomial distribution. The predicted number of failures for the primary safety endpoint will be estimated from these beta-binomial distributions. The observed and predicted number of failures will provide estimates for the anticipated study outcomes with the sample size at the time of the interim analysis and the maximum sample size allowed. The Bayesian predictive probability of success for safety for each case will be calculated from these estimates.

Enrollment will be stopped if the predictive probability of success for safety with the current sample size is greater than 90% or if the predictive probability of success for safety using the maximum sample size allowed is less than 6.5%. Otherwise, enrollment will continue until the 230<sup>th</sup> subject in the mITT population is recruited. Regardless of whether enrollment stops at one of the interim analyses or continues to a sample size of 230 subjects, the primary safety and effectiveness analyses will be conducted after all subjects have completed their 3-month post-procedure evaluation.

#### **11.4.4 Handling of Missing Data**

##### Interim Analysis:

At the time of each interim analysis, some subjects will not have completed the full 30-day follow-up for their safety evaluation. The number of failures for the primary safety endpoint in subjects with incomplete information will be assumed to follow a beta-binomial distribution. The predicted number of failures for the primary safety endpoint will be estimated from these beta-binomial distributions for estimating the predictive probability of success for the safety endpoint.

##### Final Analysis:

The reasons for missingness, if available, will be listed for subjects with missing data. Patients with missing effectiveness data will be excluded from the primary effectiveness analysis.

Subjects without full follow-up required per protocol for the safety endpoint assessment will be excluded in the primary analysis unless they have experienced a PAE. Sensitivity analyses for missing data will be performed to assess the impact of missing values on the primary safety outcome and are described in the section 11.4.5.

#### **11.4.5 Sensitivity Analyses**

To investigate the robustness of the primary safety endpoint results, the following sensitivity analyses will be performed.

##### *1) Tipping Point Analysis in mITT Population*

In this study, the tipping point is defined as the number of subjects with primary AEs for the binary safety endpoint in the study population at which the study conclusion is changed. The tipping-point analysis will replace the missing values with primary AE occurrence one at a time and the corresponding exact one-sided 95% upper confidence bound will be produced<sup>3</sup>. The confidence bound is calculated to test the PAE rate against the performance goal of 15%. If the exact one-sided 95% upper confidence bound calculated based on the imputed number of PAEs leads to a change in the study conclusion, then the imputed number of PAEs will be considered the tipping point.

The tipping-point analysis for binary safety endpoints will analyze all possible cases that may occur in the missing data cohort. The results will be presented in a summary table to illustrate how the exact one-sided 95% upper confidence bound and the conclusion on meeting the performance goal change.

##### *2) Primary Safety Analysis in Safety Population*

The primary safety analysis will be repeated in the Safety Population to include all enrolled subjects who had study catheter inserted in the evaluation of safety. The analysis is based upon data available and the early onset PAEs will be carried forward. The exact binomial proportion will be presented together with the one-sided 95% upper bound of the confidence interval.

To investigate the robustness of the primary effectiveness endpoint results, the primary effectiveness analysis will also be performed in the Safety Population for all enrolled subjects who had study catheter inserted and RF delivered with study catheters. The analysis is based upon data available. The exact binomial proportion will be presented together with the one-sided 95% lower bound of the confidence interval.

Biosense Webster, Inc

#### 11.4.6 Subgroup Analyses

Summary statistics and listings of the primary safety and effectiveness endpoints will be presented by the following factors:

- Age group: <60 vs. ≥60 years
- Repeat procedure during blanking period: Repeat procedure conducted vs. Not conducted
- Study Site
- Sex
- Operator's ablation procedure experience level: More experienced (participated in RADIANCE study or balloon experience ≥15 cases per year) vs. Less experienced (balloon experience <15 cases per year)

Subgroup analyses will be performed in the PP population for the primary effectiveness endpoint and in the mITT population for the primary safety endpoint.

#### 11.5 Analysis of Secondary Endpoints

No formal statistical hypothesis and inferential statistics will be formulated and performed for the secondary endpoints. Analyses of all secondary endpoints will be performed descriptively in the proposed analysis populations excluding the subjects with missing outcomes.

##### 11.5.1 Secondary Safety Analysis

The PAE will be summarized by the individual type of composite. Serious AEs will be summarized by the relationship to the device (SADEs) and to the procedure and by three timeframes (ablation to ≤7 days, >7 to 30 days, and >30 days post ablation). Hospitalization for cardiovascular events, which cause prolonged stay ≥2 nights post index procedure or in-patient stay not concurrent with index procedure ≥ 1 calendar day, will be summarized. Non-serious AEs will be summarized descriptively.

These analyses of secondary safety endpoints will be performed in the SP, mITT population and roll-in subjects.

##### 11.5.2 Secondary Effectiveness Analysis

###### 11.5.2.1 Secondary Effectiveness Endpoints of Effectiveness Success

Secondary effectiveness endpoints, including 6-Month Effectiveness Success and 12-Month Effectiveness Success will be summarized descriptively for all subjects in the PP population and roll-in subjects treating the endpoint as a binary variable. The point estimate and the one-sided 95% exact binomial lower bound will be presented. The secondary effectiveness analysis will



also be performed in the Safety Population for all enrolled subjects who had study catheter inserted and RF delivered with study catheters.

#### Time to Event Analysis:

Time to event analysis will also be performed for these secondary effectiveness endpoints. Kaplan-Meier curves and survival estimates will be provided using data on all subjects in the PP population and roll-in subjects. The subjects who do not have effectiveness failures but are followed up less than the endpoint-specified evaluation periods will be censored at their last observations.

#### Subgroup Analysis:

In order to provide additional characterization and interpretation of the outcomes of 6-Month and 12-Month effectiveness success, subgroup analyses based on AAD history at baseline will be performed in the PP population.

For AAD history, subjects will be categorized into AAD refractory and the AAD naïve subgroups. The AAD refractory subgroup will consist of subjects who at any point prior to their enrollment received a Class I, II, III or IV AAD for treating AF. The AAD naïve subgroup will include subjects who never received any Class I to IV AADs prior to their enrollment.

The number and percentage of subjects with the secondary effectiveness endpoints will be presented in each subgroup. Kaplan-Meier estimates will also be presented by AAD history at baseline.

#### **11.5.2.2 PVI and Non-PV Triggers Touch-Up with Focal Catheter**

The number and percentage of subjects who have pulmonary vein isolation (PVI) touch-up by focal catheter among all targeted veins will be summarized with descriptive statistics. The number and percentage of PVs with touch-up by focal catheter among all targeted veins will be summarized with descriptive statistics.

The number and percentage of subjects who undergo ablation procedure using focal catheter for non-PV triggers will be summarized with descriptive statistics.

These analyses will be performed in the PP population and roll-in subjects.

#### **11.5.3 NAE Endpoints**

The occurrences of asymptomatic and symptomatic cerebral emboli will be summarized by timepoints of pre- and post-ablation procedure. Frequency, anatomic location, and size (diameter and volume) of cerebral emboli at baseline, post-ablation and during follow-up will also be

Biosense Webster, Inc

summarized. Neurologic deficits observed post ablation and during follow up period will be summarized.

The National Institutes of Health Stroke Scale (NIHSS) scores and Montreal Cognitive Assessment (MoCA) scores at baseline, post ablation and during follow-up will be summarized with descriptive statistics.

These analyses will be performed in the NAE population.

### 11.6 Analysis of Procedural Data

Procedural data such as total procedure duration, fluoroscopy duration and dose, RF application time, balloon dwell time, time to effect (PVI), number and time of RF applications per PV location, fluid delivery, output and balance will be summarized with descriptive statistics. These analyses will be conducted using the PP population and the roll-in subjects.

### 11.7 Analysis of Atrial Fibrillation Effect on Quality-of-life Questionnaire

AFEQT includes 20 questions on a 7-point Likert scale. Questions 1 - 18 evaluate Health Related Quality of Life (HRQoL) and Questions 19- 20 relate to patients' satisfaction with treatment <sup>[4]</sup>.

First 18 questions are also used to calculate Overall AFEQT score and subscale scores across three domains

- Symptoms: Four questions 1 - 4 assess AF related symptoms
- Daily Activities: Eight questions (5 – 12)
- Treatment Concerns: Six questions (13 – 18)

The formula to calculate these scores is as follows:

$$100 - \left[ \frac{(\text{sum of severity for all questions answered} - \text{number of questions answered}) \times 100}{\text{total number questions answered} \times 6} \right]$$

Overall and subscale scores range from 0 to 100. A score of 0 corresponds to complete disability, while a score of 100 corresponds to no disability.

Baseline values and changes from baseline at each time point the questionnaire is administered will be summarized descriptively for the following five scores.

- Overall AFEQT Score (18 questions)
- Symptom Subscale Score (4 questions)
- Daily Activities Subscale Score (8 questions)
- Treatment Concern Subscale Score (6 questions)

Biosense Webster, Inc

- Treatment Satisfaction Score (2 questions)

These analyses will be conducted in the PP population and roll-in subjects.

### **11.8 Analysis of Health Economic Data**

The health economic data collected in this study includes the cost and frequency of health care utilization during hospitalization for the study index ablation procedure, as well as any additional hospitalizations during the study period. Because this data does not support the safety and effectiveness of the study catheters and the EPU, it will not be presented in the final report.

## 12 APPENDIX – ADAPTIVE DESIGN SIMULATION REPORT

# Adaptive Design Simulation Report

## 1. Study Design

This clinical investigation is a prospective, multicenter, single arm clinical evaluation utilizing the Biosense Webster multi-electrode radiofrequency balloon catheter and the Biosense Webster multi-electrode circular diagnostic catheter. The sample size for the study is primarily driven by the safety endpoint. An adaptive Bayesian design<sup>1</sup> will be used to determine the sample size based on the safety endpoint alone. The primary safety and effectiveness endpoints will be evaluated using exact tests for binomial proportions at a one-sided 5% significance level.

The null and alternative hypotheses are for primary effectiveness endpoint are:

Hypothesis  $H_0: P_E \leq 0.80$

$H_a: P_E > 0.80$

where  $P_E$  is the Proportion of patients with acute procedural success defined as confirmation of entrance block in treated PVs after adenosine and/or isoproterenol challenge. The null and alternative hypotheses for the Primary Safety Endpoint are:

Hypothesis  $H_0: P_S \geq 0.15$

$H_a: P_S < 0.15$

where  $P_S$  is the rate of Primary Adverse Events

## 2. Sample Size Justification

The study will utilize an adaptive design. The methods described in Broglio et al.<sup>[1]</sup> will be used to determine the sample size based on the safety endpoint alone. Sample size selection interim analyses will be performed when 80, 130, and 180 patients are available for the primary safety endpoint analysis. Predictive probabilities for trial success will be used to determine whether the sample size at the time of the interim analysis will be sufficient or if the trial will continue to the full sample of 230. Power for the effectiveness endpoint assessment is >80% at all sample sizes  $\geq 80$  subjects.

The final stopping sample size of 230 was determined by power calculations for the primary safety endpoint. Based on a performance goal of 15% for the primary safety endpoint and an anticipated rate of at most 9.5% for the primary safety endpoint, 230 subjects will be sufficient to obtain at least 80% power at a one-sided significance level of 5%.

It is assumed that the attrition rate for the study will be negligible. Study enrollment will continue during each interim analysis until a decision is reached to stop enrollment based on the interim

Biosense Webster, Inc

results. Due to the continued enrollment during this time, we expect an additional 5 to 6 subjects enrolled into the study. These additional subjects will be included in the primary analyses and should account for the potential subject attrition.

### 3. Adaptive Sample Size Determination

In this section, an incidence of any primary adverse events will be referred to as a failure for the primary safety endpoint. Sample size selection interim analyses will be performed when 80, 130, 180, and 230 evaluable subjects (excluding roll-in) are available for the primary safety endpoint analysis. Safety outcome at 30 days will be used as a proxy for the 3-month primary safety endpoint at each interim. The final safety analysis will be based on complete follow-up for the primary safety endpoint for all evaluable patients. Predictive probabilities of success will be used to determine whether the sample size at each interim analysis will be sufficient or if the trial enrollment will continue. Sample size simulations were performed using performance goals of 15% and 80% respectively for the safety and effectiveness endpoint rates.

At the time of each interim analysis, predictive probabilities of success for safety will be calculated once using the current sample size and again using the maximum sample size allowed. A uniform prior distribution will be assumed for the safety rate at the first interim analysis, and the distribution will be updated with the observed data at each subsequent interim analysis. The number of failures for the primary safety endpoint in subjects with incomplete information will be assumed to follow a beta-binomial distribution. The predicted number of failures for the primary safety endpoint will be estimated from these beta-binomial distributions. The observed and predicted number of failures will provide estimates for the anticipated study outcomes with the sample size at the time of the interim analysis and the maximum sample size allowed. The Bayesian predictive probability of success for safety for each case will be calculated from these estimates. No primary effectiveness endpoint analyses will occur until final sample size selection based on the primary safety endpoint has taken place.

Enrollment will be stopped if the predictive probability of success for safety with the current sample size is greater than 90% or if the predictive probability of success for safety using the maximum sample size allowed is less than 6.5%. Otherwise, enrollment will continue until the 230th subject is recruited. The Statistical Analysis Plan will contain details on all planned analyses. Regardless of whether enrollment stops at one of the interim analyses or continues to a sample size of 230 subjects, the primary analyses of the primary safety and effectiveness endpoints will be conducted after all subjects have completed their 3-month post-procedure evaluation.

### 4. Power and Type-I Error Simulations

Trial simulations were performed to estimate the Type-I error rate and the power for the success of the safety endpoint under a range of assumptions for the true safety rate and the assumed enrollment schedule. Three different enrollment schedules were considered for the simulations



corresponding to realistic, conservative, and aggressive enrollment schedules based on logistical considerations:

**Table 1 Assumed Enrollment Schedules**

	Expected number of patients per month													
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14
Realistic	1.7	3.4	5.1	8.5	11.9	17.0	20.4	23.8	25.5	25.5	25.5	25.5	25.5	25.5
Conservative	1.2	3.0	6.0	10.5	15.0	18.0	21.0	22.5	22.5	22.5	22.5	22.5	22.5	22.5
Aggressive	2.5	5.0	10.0	15.0	22.5	30.0	35.0	37.5	37.5	37.5	37.5	37.5	37.5	37.5

Table 2.a provides the estimated power as a function of the true safety rate under the realistic enrollment schedule based on 10,000 simulations for each scenario. In addition, the table provides the proportion of times we expect to stop enrollment at each interim look either for adequate sample size or for futility. Assuming a true safety rates of 0 to 9.0%, the power for showing success for the safety endpoint is estimated to be greater than or equal to 80.0%. For a safety rate of 9.5% this power decreases to approximately 73.0%. With a true effectiveness of rate of 96.0% and a performance goal of 80.0%, a minimum of 80 patients will provide 99.9% power for meeting the primary effectiveness endpoint. Therefore, the power estimates for the primary safety endpoint closely estimate the overall power of the study.

Since the primary analysis is performed at a 5.0% level Type-I error and the sample size is selected independent of the effectiveness endpoint, the Type-I error for the effectiveness hypothesis is controlled at 5.0%. To estimate the Type-I error for the primary safety endpoint, we considered the hypothetical scenario that the primary safety endpoint is on the decision boundary (highest possible type-I error, i.e. equal to 15%). Based on 25,000 trial simulations, the estimated Type-I error for this scenario was 4.7% (1,175 out of 25,000) with a 95% exact confidence interval of (0.0444, 0.0497). As the primary safety rate moves away from the decision boundary under the null hypothesis (higher than 15%), the Type-I error for the test decreases. Since our study success is defined as meeting both primary safety and effectiveness endpoints and because the Type-I error for each of these tests is controlled at 5.0%, the overall study Type-I error will be controlled at 5.0%.

**Table 2.a Estimated power as a function of true safety rate under the realistic enrollment assumption using 10,000 simulations for each scenario**

Safety rate	power	N=80	N=130	N=180	N=230
0.050	0.9959	0.4507	0.3913	0.1293	0.0287

Biosense Webster, Inc

0.055	0.9929	0.3838	0.3981	0.1664	0.0517
0.060	0.9868	0.3392	0.3776	0.2003	0.0829
0.065	0.9764	0.2856	0.3610	0.2251	0.1283
0.070	0.9630	0.2571	0.3316	0.2381	0.1732
0.075	0.9384	0.2223	0.2884	0.2437	0.2456
0.080	0.9071	0.2019	0.2678	0.2289	0.3014
0.085	0.8661	0.1881	0.2433	0.2214	0.3472
0.090	0.8005	0.1922	0.2083	0.2079	0.3916
0.095	0.7323	0.1924	0.1873	0.1952	0.4251

**Table 2.b Estimated Type-I error on the decision boundary under the realistic enrollment assumption and based on 25,000 simulations**

Safety rate	Estimated Type-I Error	95% Confidence Interval
0.15	0.0470	(0.0444, 0.0497)

Table 2.c Provides the proportion of times we would stop the trial at each interim for futility. As the true safety rate increases the likelihood of stopping at the interim analyses for futility also increases, with the highest likelihood being equal to 10.67% for stopping at with 80 patients if the true rate is 9.5%.

**Table 2.c Proportion of times the trial will stop at each interim for futility under the realistic enrollment assumption based on 10,000 simulations for each scenario**

safety rate	N=80	N=130	N=180
0.050	0.0035	0.0004	0.0001
0.055	0.0058	0.0005	0.0002
0.060	0.0104	0.0011	0.0006
0.065	0.0174	0.0024	0.0013
0.070	0.0257	0.0044	0.0022
0.075	0.0347	0.0089	0.0055

Biosense Webster, Inc

0.080	0.0475	0.0123	0.0095
0.085	0.0604	0.0210	0.0158
0.090	0.0851	0.0311	0.0243
0.095	0.1067	0.0439	0.0354

Tables 3.a, 3.c, 4.a, and 4.c provide the estimated power and futility proportions under the conservative and aggressive scenarios. The estimated type-I error assuming a safety rate of 15% was estimated to be 0.04584 (1,146 out of 25,000) with a 95% exact binomial confidence interval of (0.0432, 0.0485) under the conservative enrollment assumption (Table 3.b), and 0.04208 (1,052 out of 25,000) with a 95% exact confidence interval of (0.0396, 0.0446) under the aggressive enrollment assumption (Table 4.b).

**Table 3.a Estimated power as a function of true safety rate under the conservative enrollment assumption using 10,000 simulations for each scenario**

Safety rate	power	N=80	N=130	N=180	N=230
0.050	0.9970	0.4641	0.3996	0.1125	0.0238
0.055	0.9926	0.3972	0.4061	0.1525	0.0442
0.060	0.9870	0.3427	0.4105	0.1766	0.0702
0.065	0.9797	0.2993	0.3820	0.2034	0.1153
0.070	0.9634	0.2584	0.3567	0.2180	0.1669
0.075	0.9397	0.2281	0.3228	0.2263	0.2228
0.080	0.9114	0.2098	0.2893	0.2277	0.2732
0.085	0.8662	0.1959	0.2654	0.2145	0.3242
0.090	0.8049	0.1843	0.2208	0.2104	0.3845
0.095	0.7364	0.1917	0.2088	0.1946	0.4049

**Table 3.b Estimated Type-I error on the decision boundary under the conservative enrollment assumption and based on 25,000 simulations**

Safety rate	Estimated Type-I Error	95% Confidence Interval
-------------	------------------------	-------------------------

Biosense Webster, Inc



0.15	0.04584	(0.0432, 0.0485))
------	---------	-------------------

**Table 3.c Proportion of times the trial will stop at each interim for futility under the conservative enrollment assumption based on 10,000 simulations for each scenario**

safety rate	N=80	N=130	N=180
0.050	0.0028	0.0002	0.0000
0.055	0.0061	0.0006	0.0002
0.060	0.0093	0.0018	0.0003
0.065	0.0145	0.0021	0.0017
0.070	0.0223	0.0043	0.0035
0.075	0.0350	0.0084	0.0051
0.080	0.0462	0.0116	0.0092
0.085	0.0624	0.0223	0.0142
0.090	0.0775	0.0301	0.0255
0.095	0.1015	0.0395	0.0354

**Table 4.a Estimated power as a function of true safety rate under the aggressive enrollment assumption using 10,000 simulations for each scenario**

Safety rate	power	N=80	N=130	N=180	N=230
0.050	0.9918	0.3349	0.4029	0.1955	0.0667
0.055	0.9840	0.2836	0.3851	0.2278	0.1035
0.060	0.9793	0.2433	0.3535	0.2561	0.1471
0.065	0.9637	0.2082	0.3215	0.2594	0.2109
0.070	0.9384	0.2008	0.2695	0.2595	0.2702
0.075	0.9098	0.1784	0.2459	0.2498	0.3259
0.080	0.8714	0.1760	0.2112	0.2236	0.3892

Biosense Webster, Inc

0.085	0.8217	0.1705	0.1884	0.2103	0.4308
0.090	0.7721	0.1883	0.1677	0.1887	0.4553
0.095	0.6917	0.1960	0.1562	0.1703	0.4775

**Table 4.b Estimated Type-I error on the decision boundary under the aggressive enrollment assumption and based on 25,000 simulations**

Safety rate	Estimated Type-I Error	95% Confidence Interval
0.15	0.04208	(0.0396, 0.0446)

**Table 4.c Proportion of times the trial will stop at each interim for futility under the aggressive enrollment assumption based on 10,000 simulations for each scenario**

safety rate	N=80	N=130	N=180
0.050	0.0071	0.0010	0.0001
0.055	0.0130	0.0023	0.0005
0.060	0.0172	0.0026	0.0006
0.065	0.0270	0.0045	0.0021
0.070	0.0437	0.0093	0.0045
0.075	0.0575	0.0168	0.0069
0.080	0.0756	0.0219	0.0114
0.085	0.0980	0.0291	0.0202
0.090	0.1189	0.0350	0.0267
0.095	0.1434	0.0501	0.0381

Based on the simulation results, we concluded that the study is adequately powered to meet the primary safety and effectiveness endpoints and that the Type-I error for the overall trial success is controlled at 5% for the different enrollment schedules considered.

### 13 REFERENCES

- [1] Broglio KR, Connor JT, Berry SM. (2014) Not too big, not too small: A Goldilocks approach to sample size selection. J Biopharmaceutical Statistics. 24(3):685-705.
- [2] DerSimonian R, Laird N. (1986) Meta-analysis in clinical trials. Controlled Clinical Trials 7: 177-188.
- [3] Xu Y, Shiowjen L, Ning L. (2009) Missing Data Handling Methods in Medical Device Clinical Trials. J Biopharmaceutical Statistics. 19:1085-1098.
- [4] Atrial Fibrillation Effect on Quality-of-life (AFEQT™) Questionnaire Instruction and Scoring Manual (2009), Version 1.0, St Jude Medical Inc.