Medtronic

Clinical Investigation Plan Title	STOP AF First
Clinical Investigation Plan Identifier	MDT16012AFS001
Study Product	Arctic Front Advance™ Cardiac CryoAblation Catheter
Application Number	IDE #G1160219
ClinicalTrials.gov Identifier	NCT03118518
Sponsor	Medtronic 8200 Coral Sea St NE Mounds View, MN 55112
Document Title	STOP AF First Statistical Analysis Plan
Document Version (Date)	2.0 (09-MAY-2018)

Medtronic

Statistical Analysis Plan

Clinical Investigation Plan Title	STOP AF First
Clinical Investigation Plan Identifier	MDT16012AFS001
Sponsor/Local Sponsor	United States of America Medtronic, Inc. 8200 Coral Sea St NE Mounds View, MN U.S.A. 55112 Europe Medtronic, Bakken Research Center B.V. Endepolsdomein 5 6229 GW Maastricht The Netherlands
Document Version	2.0

Medtronic

Clinical Investigation Plan Title	STOP AF First
Clinical Investigation Plan Identifier	MDT16012AFS001
Study Product	Arctic Front Advance™ Cardiac CryoAblation Catheter
Application Number	IDE #G1160219
ClinicalTrials.gov Identifier	NCT03118518
Sponsor	Medtronic 8200 Coral Sea St NE Mounds View, MN 55112
Document Title	STOP AF First Statistical Analysis Plan
Document Version (Date)	2.0 (09-MAY-2018)

Table of Contents

1. Version History 3

2. List of Abbreviations and Definitions of Terms 3

3. Introduction 4

4. Investigation Plan 4

5. Study Objectives 6

5.1. Primary Efficacy Objective: 6

5.2. Primary Safety Objective: 6

5.3. Secondary Objectives 6

[redacted]

6. Statistical Methods 7

6.1. Study Subjects 7

6.2. General Methodology and Evaluation of Objectives 8

6.3. Center Pooling 22

6.4. Handling of Missing, Unused, and Spurious Data and Dropouts 23

6.5. Adjustments for Multiple Comparisons 25

6.6. Interim Analyses 25

6.7. Subgroup Analyses 26

6.8. Changes to Planned Analysis 26

7. Determination of Sample Size 28

7.1. Sample Size Determination for Primary Efficacy Objective 28

7.2. Sample Size Determination for Primary Safety Objective 29

8. Validation Requirements 31

9. References 32

[redacted]

1. Version History

Version	Effective Date	Summary of Changes	Change Author
1.0	07-JUN-2017	Not Applicable, New Document	Christopher Anderson, Sr. Statistician Jeff Cerkvenik, Sr. Principal Statistician
2.0	09-MAY-2018	[redacted] Changes to planned analysis section (6.8) updated by removing any references about inconsistency with the CIP in the left atrial dwell time definition. Changed '210 enrolled' to '210 randomized' where appropriate (section 6.3, sections 7.1 – 7.2), per the change in version 6 of the CIP on 16JAN18.	Christopher Anderson, Sr. Statistician

2. List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AAD	Anti-Arrhythmic Drug
AE	Adverse Event
AF	Atrial Fibrillation
AF/AFL/AT	Atrial Fibrillation, Atrial Flutter, or Atrial Tachycardia
Blanking period	The 90-day optimization period for the AAD arm, and the 90-day blanking period for the cryoablation arm.
CIP	Clinical Investigation Protocol
CTR	Cryo-Treated as Randomized
CTC	Cryo-Treated as Randomized 12-Month Completers
HCU	Healthcare Utilization
mITT	Modified Intention-To-Treat
PG	Performance Goal
PMA	Pre-Market Approval
QoL	Quality of Life
SAP	Statistical Analysis Plan

3. Introduction

The purpose of the Statistical Analysis Plan (SAP) for the STOP AF First trial is to provide pre-analysis documentation and rationale for the statistical procedures that will be employed in the planned analyses that are performed throughout this study. Specifically, this plan outlines methods used in the study's final report, as well as the primary study results publication. It does not limit the analysis that will be completed, as further analysis beyond what is specified in this document is likely.

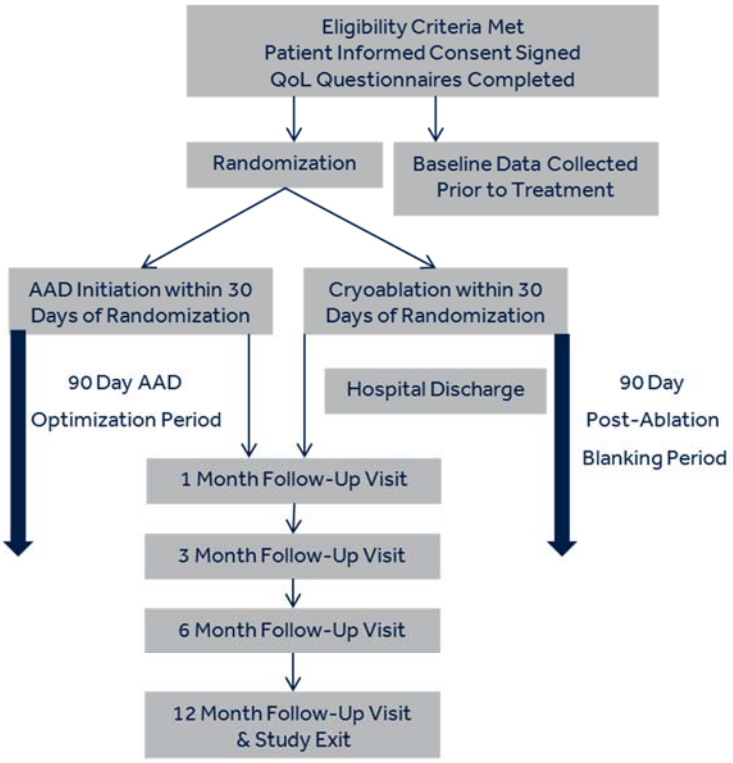
This SAP was developed based on version 6.0 of the STOP AF First Clinical Investigation Protocol (referred to as the CIP in this SAP), dated 16-JAN-2018. Topics included in this document which are not included in the CIP are handling of missing data (section 6.4), subgroup analyses (section 6.7), and validation requirements (section 8).

4. Investigation Plan

The purpose of the STOP AF First study is to provide data demonstrating the safety and effectiveness of the Arctic Front Advance™ Cardiac CryoAblation Catheter for the treatment of recurrent symptomatic paroxysmal Atrial Fibrillation (AF), without the requirement that subjects be drug refractory. The current indication is the US is as follows: The Arctic Front Advance™ Cardiac CryoAblation Catheter is indicated for the treatment of *drug refractory* recurrent symptomatic paroxysmal atrial fibrillation. The proposed indication in the US is: The Arctic Front Advance™ Cardiac CryoAblation Catheter is indicated for the treatment of recurrent symptomatic paroxysmal atrial fibrillation. The proposed indication is already within the approved indications for use in Europe.

Medtronic, Inc. is sponsoring the STOP AF First study; a prospective, interventional, multi-center, randomized, controlled, unblinded clinical study. The study will compare cryoablation to Anti-Arrhythmic Drug (AAD) therapy, the current standard of care for subjects experiencing first-time repeat occurrences of paroxysmal AF. Up to 30 sites in the US, and up to 10 in Europe will participate in the trial. Two hundred and ten (210) subjects will be randomized. Subjects will be randomized 1:1 to either AAD therapy (control arm) or cryoablation (treatment arm). Subjects will be followed at 1 month, 3 months, 6 months, and 12 months following either cryoablation or AAD initiation. Subjects will be exited from the study at the 12-month follow-up visit. Guidance for procedures completed at study visits are provided in the CIP including evaluations at baseline, cryoablation procedure, and scheduled follow up visits. No interim analyses are planned.

Figure 1: Study Design



5. Study Objectives

5.1. Primary Efficacy Objective:

- Demonstrate the superiority of cryoballoon ablation as compared to AAD therapy in terms of the rate of freedom from AF/AT/AFL (Atrial Fibrillation, Atrial Flutter, or Atrial Tachycardia) in a non-drug refractory paroxysmal AF population.

5.2. Primary Safety Objective:

- Demonstrate an acceptable safety profile of the cryoballoon ablation procedure as a first line therapy in a non-drug refractory paroxysmal AF population.

5.3. Secondary Objectives

- Assess changes in Quality of Life (QoL) between baseline and 12 months in the cryoablation arm.
- Compare Health Care Utilization (HCU) between the treatment and control arms.

6. Statistical Methods

6.1. Study Subjects

6.1.1. Disposition of Subjects

Subjects will have their eligibility assessed at baseline. The final report will provide a listing of any subjects that fail to satisfy the study's entry criteria. In the final report, a diagram or figure will describe (at a minimum) the following:

- Number of eligible subjects who consented and enrolled
- Of consented and enrolled, number randomized to cryoablation and number randomized to AAD therapy
- Subjects withdrawing consent or withdrawn by investigator prior to therapy initiation (e.g., no cryoablation despite randomization assignment due to cardiac thrombus being visible on TEE).
- Subjects with initiation of therapy as randomized.
- Withdrawals, early exits, deaths and planned study exits (i.e. after the 12-month follow-up visit) that occur after treatment initiation.

These will be described by randomized group wherever appropriate.

6.1.2. Clinical Investigation Plan (CIP) Deviations

Protocol deviations will be described using frequency tables and listings. Subjects will be excluded from analysis of primary efficacy and secondary HCU objectives if the subject did not consent to the procedure/AAD therapy initiation, or if the subject withdraws consent to be studied prior to treatment.

Handling of visit window deviations as it affects the study's primary, secondary and ancillary endpoints is discussed further in section 6.2.

6.1.3. Analysis Sets

The **Full analysis set** consists of all enrolled subjects who consent and meet the inclusion/exclusion criteria.[redacted]

The **Modified Intention-to-Treat (mITT)** cohort is the subset of subjects who maintain informed consent at least until the initiation of a cryoablation procedure or commencement of AAD treatment. For endpoints analyzed in this group of subjects, the standard intention-to-treat protocol applies immediately upon first receipt of a treatment (regardless of whether the assigned treatment was actually received). The mITT cohort will be used in the analysis of the primary efficacy objective, the secondary objectives regarding HCU and cardioversion, [redacted].

The **Cryo-Treated as Randomized (CTR)** set refers to eligible and consenting subjects who were randomized to cryoablation therapy and had a cryoablation procedure. The procedure is considered as

having been performed if a cryoballoon catheter was inserted into a subject's vasculature. This analytical cohort does not include AAD arm subjects who cross over into the cryoablation group, and does not include subjects who withdraw consent or are withdrawn by an investigator prior to their first day at risk for events. This set will be used in the analysis of the primary safety objective, [redacted].

6.2. General Methodology and Evaluation of Objectives

6.2.1. Overview

The analysis described in this SAP will be conducted by Medtronic statisticians. Prior to evaluation of the study's primary objectives, a descriptive analysis will be performed: demographic and other key baseline characteristics will be summarized by randomization group for the mITT dataset. Additional exploratory analyses of the data may be conducted as deemed appropriate.

6.2.2. Primary Efficacy Objective

Demonstrate the superiority of cryoballoon ablation as compared to AAD therapy in terms of the rate of freedom from AF/AT/AFL in a non-drug refractory paroxysmal AF population.

6.2.2.1. Hypothesis

The hypothesis test used in assessment of the primary efficacy objective is that the proportion of subjects with treatment success at 12 months is greater in subjects using cryoballoon ablation compared to subjects using AAD therapy.

The following hypothesis will be tested in a two-sided test with $\alpha = 0.05$:

H_0: $\pi_{cryo} = \pi_{AAD}$

H_a: $\pi_{cryo} \neq \pi_{AAD}$

Where π_{cryo} and π_{AAD} are the proportion of treatment successes at 12 months in the modified intention-to-treat cohorts of the cryoballoon ablation and AAD groups, respectively.

6.2.2.2. Endpoint Definition

A treatment success is the opposite of a treatment failure.

A subject is considered a treatment failure if he/she experiences any of the following:

- Acute procedural failure (treatment arm only), which is any of these:
  - Inability to isolate all accessible targeted pulmonary veins (assessed for entrance block and, where assessable, exit block) during the index ablation procedure, OR
  - Left atrial non-PVI ablations including but not limited to, ablation of linear lesions, OR
  - Use of a non-study device in the left atrium.
- Documented AF/AT/AFL on ambulatory monitoring/12-lead ECG after the 90-day post-ablation blanking period (treatment arm).
  - Minimum of 30 seconds on ambulatory monitoring or 10 seconds on 12-lead ECG.
  - Note: Documented occurrence and treatment of typical right-sided cavotricuspid isthmus dependent atrial flutter is not considered a failure if confirmed by entrainment maneuvers during EP testing.
- Any subsequent AF surgery or ablation in the left atrium.
- Any subsequent cardioversion after the 90-day post-ablation blanking period.
- Class I or III antiarrhythmic drug (or sotalol) use after the 90-day blanking period (treatment arm only).

The blanking period is defined as the first 90 days after the index ablation procedure (treatment arm) or AAD initiation (control arm). Recurrences of atrial arrhythmias during the blanking period will not be counted in the determination of the first clinical failure for the primary endpoint. Within the 90-day blanking period, recurrent arrhythmias can be managed with medications or cardioversions.

6.2.2.3. Performance Requirements

This hypothesis test is for superiority of the cryoablation procedure over AAD's as a treatment for AF/AFL/AT in a drug-naïve population. If a two-sided log-rank test shows the difference in 12-month success rates to be less than the pre-specified alpha of 0.05, the null hypothesis (that cryoballoon ablation has similar 12-month efficacy rates compared to AAD therapy) will be rejected in favor of the alternative hypothesis (that 12-month success rates for subjects treated with cryoballoon ablation is greater than for those treated with AAD's).

6.2.2.4. Analysis Methods

The probability of a subject achieving success at 12 months (365 days) will be estimated using Kaplan-Meier survival analysis. The standard error will be approximated using Greenwood's formula. A two-sided 95% log-log confidence interval for the probability will be constructed. A log rank test with $\alpha = 0.05$ will be used to assess whether the failure rate differs between treatment groups.

Day 0 is defined as either the day of the index cryoablation procedure or the day in which AAD drug therapy is initiated. For subjects with treatment failure, the survival date will be set to the date of the treatment failure. For subjects without treatment failure through 12 months, those subjects will be censored at the last study contact date recorded on a CRF, which may include the last study visit, the exit date, or death date. If a subject without a treatment failure is lost to follow-up, the censoring date will be set to the last recorded study visit date. If treatment failure occurs on the day of index treatment, survival time will be set to 0.5 days.

If an occurrence of documented AF/AT/AFL resulted from rhythm monitoring that was initiated at the 12-month visit, then the date of occurrence used in the Kaplan-Meier analysis will be the minimum of 365

days from the cryoballoon ablation procedure/date of AAD initiation, or the actual date of occurrence. This allows counting of all AF/AT/ AFL events found at the 12-month follow-up, even if the follow-up occurred after 365 days post-ablation but still within the visit window. Events found at 12-month visits outside of the window will be censored at the previous follow-up and described in deviation listings.

The survival curves from 0 to 12 months will be presented for both treatment arms.

6.2.2.5. Determination of Subjects for Analysis

The mITT cohort will be used for this analysis.

6.2.3. Primary Safety Objective

Demonstrate an acceptable safety profile of the cryoballoon ablation procedure as a first line therapy in a non-drug refractory paroxysmal AF population.

6.2.3.1. Hypothesis

The following hypothesis will be tested in a one-sided test with $\alpha = 0.025$:

H_0: $P_s \geq 12\%$

H_a: $P_s < 12\%$,

where P_s is the probability of a safety event in subjects from the cryoablation arm.

6.2.3.2. Endpoint Definition

A primary safety event is defined as a serious procedure-related or cryoablation system-related adverse event that includes any of the following:

- TIA within 7 days
- Cerebrovascular accident within 7 days
- Major bleeding that requires transfusion or results in a 20% or greater fall in hematocrit (HCT) within 7 days
- Development of a significant pericardial effusion within 30 days. A significant pericardial effusion is one that results in hemodynamic compromise, requires elective or urgent pericardiocentesis, or results in a 1-cm or more pericardial effusion as documented by echocardiography.
- Symptomatic PV stenosis within 12 months; accompanied by one of the following: 50%-70% reduction in area of the pulmonary vein, with symptoms not explained by other conditions; OR >70% reduction in area of the pulmonary vein
- MI within 7 days
- PNI unresolved at 12 months
- AE fistula within 12 months
- Major vascular complication that requires intervention, prolongs the hospital stay, or requires hospital admission (within 7 days).

6.2.3.3. Performance Requirements

If the upper bound of the two-sided 95% confidence interval of the safety event rate in cryoablation arm subjects is <12%, this objective will be considered met.

6.2.3.4. Analysis Methods

The probability of a safety event within 12 months will be estimated using Kaplan-Meier survival analysis. Greenwood's formula will be used to approximate the standard error of the survival curve, and a two-sided log-log confidence interval at 12 months will be reported.

Day 0 is defined as the day of the index cryoablation procedure. For subjects with a safety event, the survival date will be set to the date of the safety event. For subjects without a safety event, censoring will occur at the last study contact date with a corresponding CRF (i.e., the last study visit, exit date or death). If a subject without a safety event is lost to follow-up they will be censored on the date of the last study visit.

Any subsequent ablations a subject may have after day 0 will not reset their survival time; the date of the index procedure will remain as day 0 for purposes of the primary safety analysis. Safety events related to repeat ablation procedures occurring up to 365 days after the index cryoablation date will be considered as having met criteria for the primary safety endpoint, and will be counted as failures for this analysis. Safety events only observed at 12-month follow-up visits that occur 365 - 395 days after the index ablation (i.e. are in the visit window) will also count as safety failures at one year. Like with the primary efficacy objective, safety events found at 12-month visits outside of the window (>395 days from index ablation) will not be counted in this endpoint, but will be described in AE listings.

6.2.3.5. Determination of Subjects for Analysis

The primary safety objective will be assessed in the CTR cohort.

6.2.3.6. Additional Safety Analysis Specifications

Subjects randomized to the AAD arm who receive a cryoablation procedure during follow-up will not be included in the primary safety analysis (as they are not included in the CTR dataset). However, events corresponding to the primary safety endpoint (treatment success, treatment safety, QoL and HCU) occurring in this cohort of subjects will be reported and summarized as well. Kaplan-Meier methods will be used to estimate event rates and corresponding confidence intervals, with day 0 being set to the date of the crossover ablation. Standard error will be approximated with Greenwood's formula, and a two-sided log-log confidence interval will be presented. No hypothesis testing is planned for analysis of crossover ablation subjects, and there is no planned comparison of this cohort to the CTR cohort. Depending on the number of repeat ablations observed over the course of the study, a similar set of analyses may also be performed on the subset of subjects having a second ablation, with day 0 being set to the date of the repeat ablation. If 10 or fewer repeat ablations are observed, associated characteristics and outcomes of repeat ablation will be described through listings.

6.2.4. Secondary Objective 1: Quality of Life

Assess changes in quality of life between baseline and 12 months in the cryoballoon ablation arm.

There are two hypotheses tested in the objective, with separate hypothesis tests for (1) the difference in composite scores from the AFEQT questionnaire taken at baseline and 12 month visits, and (2) for the difference in composite scores for the EQ-5D questionnaire taken at baseline and 12-month visits. Hypothesis tests for this objective and secondary objective 2 will have type I error controlled using the Hochberg procedure, which is described in section 6.5. Testing for secondary objectives will only be performed if the primary objectives are met.

Comparison to the AAD arm is not planned as it is expected that a large percentage of AAD arm subjects will have a cryoballoon procedure during the 12 months of follow-up. An intention-to-treat comparison between the treatment and control arms would be biased toward the null hypothesis of no difference in such circumstances.

6.2.4.1. AFEQT

6.2.4.1.1. Hypothesis

The following hypothesis will be tested in a two-sided test:

Ho: $AFEQT_{month12} = AFEQT_{baseline}$

Ha: $AFEQT_{month12} \neq AFEQT_{baseline}$,

Where $AFEQT_{baseline}$ is the score from the AFEQT assessed at the baseline visit, and $AFEQT_{month12}$ is the composite score from the AFEQT assessment from the 12-month follow-up.

6.2.4.1.2. Endpoint Definition

The AFEQT questionnaire will be utilized for this objective. The questionnaire is an atrial fibrillation specific health-related quality of life questionnaire to assess the impact of AF on a subject's life. The overall score ranges from 0 – 100, with 0 corresponding to complete disability and 100 corresponding to no disability. The absolute difference between each subject's baseline and 12 month scores is the endpoint of interest.

6.2.4.1.3. Analysis Methods

A paired t-test will be used to assess the difference in mean AFEQT scores between baseline and 12 month measurements. If the p-value from this test is less than the alpha level determined by the Hochberg procedure (described in section 0), the objective will be considered met. This analysis requires subjects to have assessments at both the baseline and annual follow-up visit. It will be performed in subjects from the CTR cohort who have paired data.

In addition to the composite score, which is of primary interest for this objective, the AFEQT questionnaire has three subscale scores: the Daily Activities Subscale, Treatment Concern, and Treatment satisfaction. Each subscale ranges from 0 – 100, where 0 corresponds to low quality-of-life and 100 corresponds to high quality of life. Change in AFEQT subscale score is defined as 12-month AFEQT subscale score minus baseline AFEQT subscale score. A two-sided 95% confidence interval will be calculated for each change in subscale score based on the t-distribution.

6.2.4.1.4. Determination of Subjects for Analysis

This analysis will be performed in the CTR cohort, but subjects who do not have paired AFEQT assessment data will be excluded.

6.2.4.2. EQ-5D

6.2.4.2.1 Hypothesis

The following hypothesis will be tested in a two-sided test:

Ho: $EQ\text{-}5D_{month12} = EQ\text{-}5D_{AAD}$

Ha: $EQ\text{-}5D_{month12} \neq EQ\text{-}5D_{AAD}$

Where $EQ\text{-}5D_{month12}$ is the mean composite EQ5D score from subjects assessed at the 12 month visit, and $EQ\text{-}5D_{baseline}$ is the mean composite EQ-5D score for at baseline.

6.2.4.2.2. Endpoint Definition

The Euroqol EQ-5D questionnaire (which consists of a 5-question survey and a visual analog scale indicating the subject's overall health) will be utilized for this objective. Composite scores will be indexed against a US reference population (Shaw, Johnson and Coons 2004). We provide SAS code that may be used as a starting point for calculating the composite EQ-5D score in appendix 10.1.

6.2.4.2.3. Analysis Methods

Change in EQ-5D composite score is defined as 12-month EQ-5D score minus baseline EQ-5D score. Differences in mean EQ-5D scores between visits will be assessed utilizing a paired t-test. A two-sided 95% confidence interval will be calculated based on the t-distribution. If the p-value from the paired t-

test is less than the alpha level determined by the Hochberg procedure (described in section 6.5), the objective will be considered met. This analysis requires subjects to have assessments at both the baseline and annual follow-up visit.

6.2.4.2.4. Determination of Subjects for Analysis

This analysis will take place in the CTR dataset, but will be limited to subjects with paired EQ-5D assessments.

6.2.5. Secondary Objective 2: Healthcare Utilization

Compare health care utilization between the treatment and control arms. There are two hypotheses tested in the objective, with separate hypothesis tests for: (1) the rate of total health care utilization events (cardiovascular-related hospitalizations, emergency room visits, or unscheduled office visits) over 12 months, and (2) the rate of cardioversions (electrical or pharmacological) over 12 months. Hypothesis tests for this objective and secondary objective 1 will have type I error controlled using the Hochberg procedure, which is described in section 6.5. Testing for secondary objectives will only be performed if the primary objectives are met.

6.2.5.1. Cardiovascular hospitalizations, ED and unscheduled office visits

6.2.5.1.1. Hypothesis

The following hypothesis will be tested.

H_0: $\theta_{cryo} = \theta_{AAD}$

H_a: $\theta_{cryo} \neq \theta_{AAD}$,

where θ_{cryo} and θ_{AAD} are the 12-month rates of cardiovascular HCU events in the cryoballoon ablation and AAD arms, respectively.

6.2.5.1.2. Endpoint Definition

Healthcare utilization (HCU) events for this objective are defined as cardiovascular-related hospitalizations, cardiovascular-related emergency department visits, or cardiovascular-related unscheduled office visits occurring within 12 months after the index ablation (cryo arm) or the initiation of therapy (AAD arm). Events occurring >365 days after initiation of treatment but before the 12-month visit (if within its window) will be included.

6.2.5.1.3. Analysis Methods

The probability of a subject achieving freedom from HCU events (defined above) at 12 months will be estimated using Kaplan-Meier survival analysis. The standard error will be approximated using Greenwood's formula. A two-sided 95% log-log confidence interval for the probability at 12 months will be constructed. A two-sided log rank test will be used to assess whether the HCU event rate differs between treatment groups. If the p-value from this log-rank test is less than the alpha level determined by the Hochberg procedure (described in section 6.5), the objective will be considered met.

Day 0 is defined as day of the index cryoablation procedure or the day in which drug therapy is initiated. For subjects with treatment failure, the survival date will be set to the date of the first HCU event. For subjects without any HCU events through 12 months, those subjects will be censored at the last study contact date recorded on a CRF which may include the last study visit, the exit date or the date of death. If a subject without any HCU events is lost to follow-up, the censoring date will be set to the last known study visit date. For HCU events occurring on the day of index treatment, survival time will be set to 0.5 days.

The survival curve from 0 to 12 months will be presented for both treatment arms. HCU events will also be summarized by type and treatment group.

6.2.5.1.4. Determination of Subjects for Analysis

This analysis will be performed in the mITT dataset.

6.2.5.2. Cardioversions

6.2.5.2.1. Hypothesis

The following hypothesis will be tested with a two-sided test.

H_0: $\gamma_{cryo} = \gamma_{AAD}$

H_a: $\gamma_{cryo} \neq \gamma_{AAD}$,

where γ_{cryo} and γ_{AAD} are the 12-month rates of cardioversion (either electrical or pharmacological) in the cryoballoon ablation and AAD arms, respectively.

6.2.5.2.2. Endpoint Definition

A cardioversion event is defined as an electrical or pharmacological cardioversion post index ablation discharge for the treatment arm, and post AAD initiation for the AAD arm.

6.2.5.2.3. Analysis Methods

The probability of a subject achieving freedom from cardioversion (defined above) at 12 months (365 days) will be estimated using Kaplan-Meier survival analysis. The standard error will be approximated using Greenwood's formula. A two-sided 95% log-log confidence interval for the probability at 12 months will be constructed. A two-sided log rank test will be used to assess whether the cardioversion rate differs between treatment groups. If the p-value from the log-rank test is less than the alpha level determined by the Hochberg procedure (described in section 6.5), the objective will be considered met.

Day 0 is defined as the day of the index cryoablation procedure or the day in which drug therapy is initiated. For subjects with at least one cardioversion, the survival date will be set to the date of the first cardioversion. For subjects without any cardioversions through 12 months, those subjects will be censored at the last study contact date recorded on a CRF which may include the last study visit, the exit

date, or the date of death. If a subject without any cardioversions is lost to follow-up, the censoring date will be set to the last known study visit date. For cardioversions occurring on the day of index treatment, survival time will be set to 0.5 days.

The survival curve from 0 to 12 months will be presented for both treatment arms. Counts of all cardioversions administered over 12 months will also be summarized.

6.2.5.2.4. Determination of Subjects for Analysis

This analysis will be performed on data from the mITT dataset.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.3. Center Pooling

The STOP AF First study is expected to be conducted at up to 10 sites in Europe and 30 sites in the US, with a total of 210 subjects randomized across all study centers. Per the CIP, each site will enroll between 0 and 31 subjects. The protocol also specifies that no more than 50% of the total enrollment will come from outside of the US.

Data from all sites will be combined without regard to center location for the analysis of the study's primary, secondary [redacted] endpoints. However, it will be assessed whether site-to-site and geographical heterogeneity exists in the rate of pre-specified primary endpoints using a fixed-effects meta-analytic approach. This analysis will investigate two things. First, whether sites exhibit significant heterogeneity in event rates, and second, whether geography (a binary variable representing whether a site is located in the US or Europe) moderates any statistically significant heterogeneity that is observed. Models will be fit separately for each primary outcome; in the analysis of each, any center having ≤5 subjects with complete data will be combined into 'US_Small' and 'EU_Small' analysis groups. If a Cochran's Q-test[1] for heterogeneity shows $p<0.05$, it will be taken as evidence of significant heterogeneity between sites. Evidence of between-site heterogeneity will not necessarily preclude pooling data; rather, it will prompt further investigation into the sources of the apparent differences in event rates between sites[2]. At a minimum, findings of analyses on heterogeneity between study sites and between all US and all European sites will be shown in the final report in a table by endpoint.

[1] Cochran, WG. "The Comparison of Percentages in Matched Samples". *Biometrika*, 1950: 37 (3-4), 256-266.

[2] Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. 2009. *Introduction to meta-analysis*, Chichester, UK: Wiley.

6.4. Handling of Missing, Unused, and Spurious Data and Dropouts

6.4.1. Tipping Point Analysis

It is expected that 10% of the subjects enrolled in this trial will exit the study before their 12-month visit. Two general strategies will be used to mitigate the impact of missing data in this study. First, most objectives will be analyzed using Kaplan-Meier survival analysis methods, which allow data from subjects lost to follow-up to still be utilized up until their last date of contact. Second, to test the sensitivity of the primary efficacy and safety analyses to the range of values possible, but unobserved in subjects exiting prior to 12 months, a tipping point analysis will be performed on the full analysis set.

Primary objective data will be sorted by ascending event times, as shown in the fictional example dataset below from an efficacy analysis where $p < 0.05$:

Treatment	Time	Event
Cryoablation	0.5	1
…		
AAD	363	1
Cryoablation	364	0
Cryoablation	365	1
AAD	371	1
…		
Cryoablation	378	0
Cryoablation	379	0
AAD	380	0

In this example, 'Treatment' is the group to which the subject was randomized, 'Time' is the number of days from treatment initiation (either the index ablation or AADs) to the 12-month follow-up, end of study, or endpoint event, and 'Event' takes a value of 0 to indicate censoring or a value of 1 to indicate that the endpoint was met. Then starting from the first time < 0, the following algorithm is followed:

1) If a failure is observed, skip steps 2 and 3;
2) If the observation is censored:
    - set the time = 365 if the observation was from an AAD subject
    - set time = time + 1 and event = 1 if the observation was from a cryoablation subject.

3) Apply the log-rank test specified in section 6.2.2.4 to find the new two-sided p-value
4) If P is still < 0.05 and a larger event time < 365 exists, repeat the process with that observation; otherwise stop.

Essentially, this process goes through the hypothetical exercise of sequentially assuming unobserved control arm subjects are successes, and unobserved treatment arm subjects are failures the day after being lost to follow-up until the log-rank test supports a different conclusion (in this case, when p >=0.05). This first subject for which p <0.05 is no longer true is the point at which the analysis 'tips'. Larger numbers of treatment failures and control successes needed to change the conclusion indicate findings that are more robust to informative censoring. If the primary efficacy analysis as specified in section 6.2.2.4 shows p ≥0.05, the tipping point analysis will be modified to quantify how many treatment arm successes and control arm failures would be necessary to reach a different conclusion for the hypothesis test.

The primary safety endpoint is only measured in the cryoablation arm. As such, the algorithm to find the tipping point of the primary safety algorithm is slightly different. Like with the efficacy tipping point analysis, data will be sorted by ascending event times, and subjects who are censored prior to 365 days are assigned an event time equal to (date of last contact –index ablation date). Then, for each observation:

1) If a failure is observed, skip steps 2 and 3;
2) If the observation is censored, set time = time + 1 and event = 1.
3) Apply the log-rank test specified in section 6.2.3.4 to find the new two-sided p-value and 95% log-log confidence interval.
4) If P is still < 0.05 and a larger event time < 365 exists, repeat the process with that observation; otherwise stop.

The first observation where the confidence interval includes 12.0% is the tipping point for the primary safety analysis. This analysis will be performed in the CTR dataset. For either primary endpoint, if fewer than 5 subjects have missing data, a worst-case analysis will be done instead of a tipping point analysis.

6.5. Adjustments for Multiple Comparisons

A Hochberg multiple testing procedure will be utilized to maintain an overall type I error rate of 0.05 for the four hypotheses being tested among the two secondary objectives. Testing for these objectives will be performed if the primary objectives are met.

The Hochberg procedure is a stepwise procedure and will be implemented as follows. The four hypotheses are abbreviated as H(1), H(2), H(3), and H(4). For each of these, p-values will be calculated and sorted $p(1) < p(2) < p(3) < p(4)$. The decision rule to accept or reject each hypothesis will follow the step-up algorithm, where $\alpha=0.05$:

Step 1: If $p(4) \geq \alpha$, accept H(4) and go to Step 2, otherwise reject all hypotheses and stop
Step 2: If $p(3) \geq \alpha/2$, accept H(3) and go to Step 3, otherwise reject H(3), H(2) and H(1) and stop
Step 3: if $p(2) \geq \alpha/3$, accept H(2) and go to Step 4, otherwise reject H(2) and H(1) and stop
Step 4: If $p(1) < \alpha/4$, reject H(1); otherwise accept H(1)

In practice, this adjustment may be done simultaneously on a dataset from the four p-values produced, [redacted]

[redacted]

6.6. Interim Analyses

No interim analyses are planned.

6.7. Subgroup Analyses

A limited number of additional analyses will be performed to evaluate evidence for a differential effect of cryoablation vs. AAD therapy on the primary and secondary endpoints within subgroups of subjects.

At the time of this SAP, current FDA guidance[3] recommends additional evaluation of primary objectives within the following demographic subgroups:

- Age (calculated as [year of randomization date – year of birth])
- Gender (captured as male or female by the CRF, with a third level for no response)
- Race (White, Black or African American, Asian, American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, Other, Not Stated)
- Ethnicity (Hispanic or Latino, or not)

Subgroup analyses will be performed separately for each of the demographic variables listed above, using Kaplan-Meier survival methods. Handling of censoring, visit windows, etc. will be consistent with what is described in the corresponding general methods above in section 6.2 for each objective. [redacted]

Demographic strata with <5 events will be combined with other strata if the resulting combination is still deemed analytically meaningful, or may be ignored due to the sparseness of information. Age will be calculated as [year of randomization date – year of birth] as the case report form does not collect the exact date of birth for a subject. For subgroup analysis, age will be divided into quartiles. If the overall log-rank test of equality over strata is significant for strata with more than two levels, Tukey's range test will be used to adjust type I error for the comparison between multiple subgroups. Subgroup analyses will be performed in the same datasets in which the corresponding primary objectives are analyzed.

6.8. Changes to Planned Analysis

Additional details on analysis methods have been added to this SAP, but only [redacted] minor changes to the statistical methods defined in the STOP AF First CIP version 4 are noted. First, the definition of the PV stenosis component of this study's primary safety endpoint has been altered according to FDA study design considerations. The criteria of 70% PV area reduction stated in this document reflects this update, whereas the CIP version 4 shows the original definition of 75% reduction. Second, the sample size calculation for the primary efficacy analysis incorrectly states that Fisher's exact test was used in the

3 https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDASIA/UCM365544.pdf

https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDASIA/UCM365544.pdf

calculation. In actuality, a Pearson's chi-squared test was used, as stated in section 7.1 of this SAP, which led to the stated analytic sample size of 174. The significance of this edit is negligible, since in both cases the planned endpoint analysis involves using Kaplan-Meier methods, and the sample size for the study is driven by the primary safety endpoint. [redacted] Analytical deviations from procedures this SAP may be addressed by the release of newer SAP versions (if feasible), or will be described in the final report and/or main manuscript, along with the rationale for the deviation.

7. Determination of Sample Size

Sample size for STOP AF First was determined by finding the minimum number of subjects enrolled which provides adequate power under the assumptions to both the hypothesis test for the primary efficacy objective, as well as the hypothesis test for the primary safety objective. Calculations are provided separately for each primary endpoint.

7.1. Sample Size Determination for Primary Efficacy Objective

Sample size was estimated using Proc Power in SAS v9.4. The calculation was done to find the number of analyzed subjects necessary to provide a two-sided Pearson's chi-square test with α = 0.05 and 90% power to detect a difference between two binomial proportions, one of which was 69.9%, and the other 45.9%, under a 1:1 randomization ratio. The two groups' rates represent the cryoablation and AAD arm 12-month success rates, respectively. The assumed cryoablation arm success rate was what was observed in the STOP AF pivotal trial. The AAD arm success rate estimate, 28 of 61 successes at one year, was observed under similar circumstances in the RAAFT-2 trial (Morillo, et al.). [redacted]

[redacted]

Computed N Total

Actual Power	N Total
0.901	174

This shows an analytical sample size of 174 subjects is required. However, it was assumed that 10% attrition would occur between enrollment and the 12-month follow-up. Therefore, the necessary enrollment calculated to yield a 90.1% probability of detecting the presumed treatment success difference is 194 subjects.

The sample size estimate attained based on using a chi-square test is conservative compared to the actual number of subjects required using the Kaplan-Meier methods described in 6.2.2, since in the latter case, attrition is partial: data is included for all subjects up until they exit the study, so only subjects who exit prior to day 0 (index ablation or AAD initiation) will provide no information to the analysis.

7.2. Sample Size Determination for Primary Safety Objective

Sample size was estimated using Proc Power in SAS 9.4. The parameters of the calculation were for a one-sided exact binomial test with α = 0.025 and 80% power, with a true safety event rate of 4.0% and a performance goal of 12%. The same conditions for attrition (10% from enrollment to 12 months) and randomization ratio (1:1) were assumed in this calculation as they were for the primary efficacy endpoint. The event rate of 4% was derived by applying the definition of the primary safety endpoint retrospectively to subjects from the STOP AF pivotal study; when this was performed, event rates of 5/163 (3.1%) or 7/163 (4.3%) were observed, depending on how asymptomatic phrenic nerve injuries were counted. [redacted]

[redacted]

Computed Power

Lower Crit Val	Upper Crit Val	Actual Alpha	Power
5	.	0.0246	0.825

Under the above conditions, 80% power is achieved with 94 cryoablation arm subjects analyzed. Assuming 10% attrition, then 105 enrolled subjects are needed, requiring 210 subjects to be enrolled and randomized 1:1.

The required sample size of 210 is the smallest number that results in adequate power for both the primary efficacy and safety analyses under the specified conditions, and will be used as the enrollment target for the study. Figures 2 and 3 (below) show the sensitivity of the STOP AF First power calculations to deviations from the event rate assumptions used above, for the efficacy and safety endpoints, respectively.

Figure 2: Efficacy Objective Power by True Cryo and AAD Success Rates

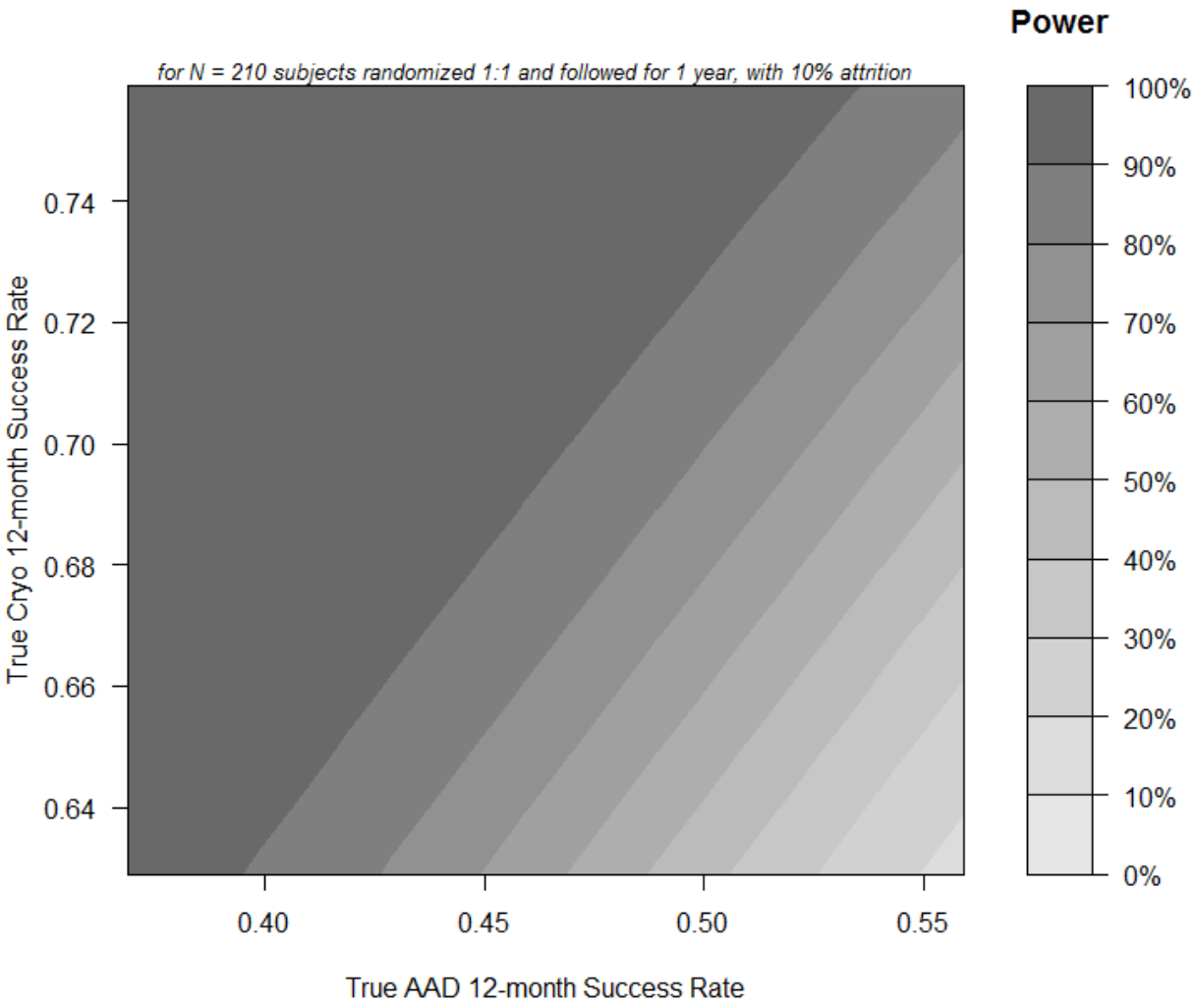
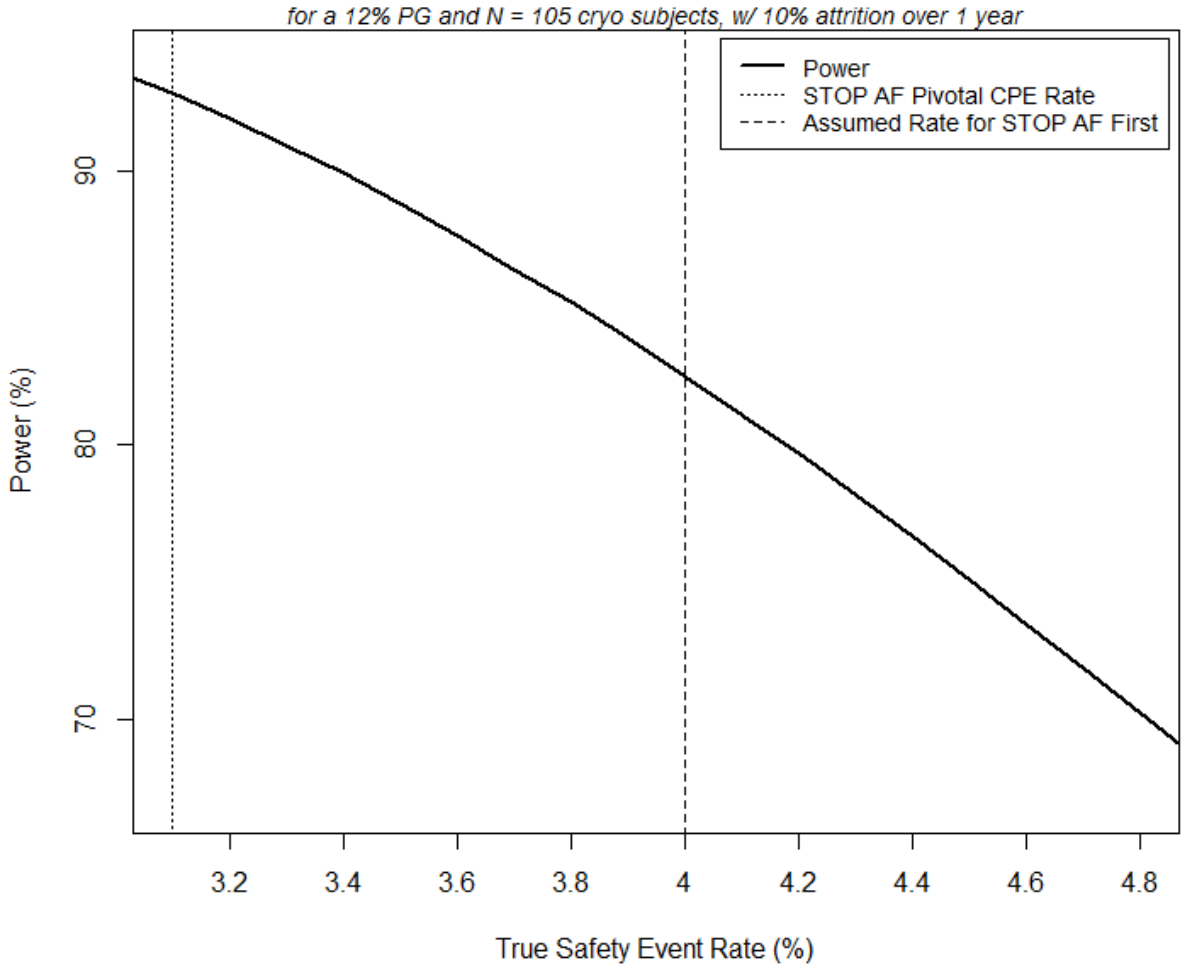


Figure 3: Safety Objective Power by True Event Rate



8. Validation Requirements

Verification of analyses of both the primary efficacy objective and the primary safety objective will be completed with level I validation (independent programming). Secondary [redacted] objectives will be validated with a minimum of level II validation. Analyses that are not related to primary objectives [redacted] will be validated at a minimum of level II validation if being presented externally in an abstract or publication.

9. References

Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. 2009. *Introduction to meta-analysis*, 107- 125, Chichester, UK: Wiley.

Cochran, WG. "The Comparison of Percentages in Matched Samples". *Biometrika*, 1950: 37 (3-4), 256-266.

Hochberg, Y. "A Sharper Bonferroni Procedure for Multiple Significance Testing," *Biometrika*, 1988: 75, 800–803.

Shaw, JW, JA Johnson, and SJ Coons. "U.S. Valuation of the EQ-5D Heatlh States: Development and Testing of the D1 Valuation Model." *Medical Care*, 2004.

Spertus, J, et al. "Development and validation of the Atrial Fibrillation Effect on Quality-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation." *Circ Arrhythm Electrophysiol*, Feb 4, 2011: 15-25.

U.S. Food and Drug Administration. "FDA Report: Collection, Analysis and Availability of Demographic Subgroup Data for FDA-Approved Medical Products." FDA Web Site. https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDASIA/UCM365544.pdf. Published August 2013. Accessed April 28, 2017.

U.S. Food and Drug Administration. "Evaluation of Sex-Specific Data in Medical Device Clinical Studies". FDA Web Site. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM283707.pdf. Published August 22, 2014. Accessed May 11, 2017.

[redacted]

[redacted]

[redacted]

[redacted]