Medtronic

Statistical Analysis Plan

Clinical Investigation Plan Title	STOP Persistent AF Clinical Investigation Plan
Sponsor/Local Sponsor	United States of America Medtronic, Inc. 8200 Coral Sea Street NE Mounds View, MN 55112 United States of America Canada Medtronic of Canada, Ltd. 99 Hereford Street Brampton, Ontario, L6Y 0R3 Canada Japan Medtronic Japan Co. Ltd. 1-2-70 Konan Minato-ku, Tokyo 108-0075 Japan
Document Version	SAP Version 4 08-JUL-19

Table of Contents

1. Version History

Version	Date	Summary of Changes	Author(s)/Title
1.0	09-JUN-2017	Version 1 of the Statistical Analysis Plan (SAP) is based off Clinical Investigational Plan Version 5. Note, CIP was updated to version 6 but no statistical methods were changed between CIP version 5 and 6.	Christopher Anderson, Sr. Statistician, CRHF Fred Kueffer, Sr. Principal Statistician, CRHF
2.0	07-NOV-2018	Version 2 includes one change. In the Secondary Objective analysis methods, the following sentence was added "H(1), H(2) and H(3) denote the null hypothesis for change in AFEQT, change in SF-12 mental component, and SF-12 physical component.". This sentence was added based on FDA Study design consideration from FDA STOP Persistent AF IDE approval letter dated November 11, 2016.	Fred Kueffer, Sr. Principal Statistician, CRHF
3.0	15-NOV-2018	Version 3 of the SAP defines an interim analysis. The addition of the interim analysis was suggested by the FDA in feedback on Q181709 dated October 17, 2018. The interim analysis methods have been added to Appendix A. An interim analysis was not previously included in SAP version 1 and 2 or the study protocol. Interim primary efficacy and primary safety objectives and statistical performance criteria have been defined. The interim analysis plan also includes all secondary and ancillary objectives for the study. Due to interim follow-up, some of the objectives have been modified to account for interim this (ie. not all subjects will have complete 12 month data). The intent of the version is to pre-defined all objectives and analyses (interim primary efficacy and primary safety, secondary and ancillary) that will be reviewed during an interim analysis. The timing of this interim analysis will occur when at least 70% of follow-up has occurred. This timing is projected to coincide with the projected last 6-month visit date. It is estimated that 75% of the total expected follow-up will be available at the timing of this analysis. Appendix A has been added and includes all the details of the interim analysis. Additional changes include three new defined subgroups: 1. Subjects with symptomatic persistent AF documented by two documented ECGs versus subjects with symptomatic persistent AF documented by one ECG and a doctor note indicating the patient had symptoms consistent with AF 2. Subjects with body mass index (BMI) > 40 versus ≤ 40.	Fred Kueffer, Sr. Principal Statistician, CRHF

Version	Date	Summary of Changes	Author(s)/Title
		Both subgroups have been added to assess changes in inclusion and exclusion criteria between clinical investigational plan version 5 to version 6. 3. Subjects with baseline ECG performed within 30 days of index ablation versus > 30 days. This subgroup has been added to address the FDA study design consideration in FDA approval letter dated January 10, 2018.	
4.0	08-JUL-2019	Removed the interim analysis defined in version 3 as well as Appendix A, which specifies the interim analysis. Reverted back to version 2, except for the additional three pre-specified subgroup analyses defined in version 3, per FDA feedback. Per FDA feedback, added a sensitivity analysis (section 6.2.4.3) to the secondary objective to assess the effect of missing data on QoL outcomes. Added additional clarfications about subgroup and pooling analyses in section 6.3. Removed references to European study centers.	Christopher Anderson, Principal Statistician, CRHF

2. List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AE	Adverse Event
AF	Atrial Fibrillation
AF/AFL/AT	Atrial Fibrillation, Atrial Flutter or Atrial Tachycardia
AFEQT	Atrial Fibrillation Effect on QualiTy-of-life assessment
CBA	CryoBalloon Ablation
CIP	Clinical Investigation Protocol
LTFU	Lost To Follow-Up; alternatively, Loss To Follow-Up
MedDRA	Medical Dictionary for Regulatory Activities
PMA	Pre-Market Approval
PMA-S	Pre-Market Approval Supplement
PMDA	Pharmaceuticals and Medical Devices Agency
PNI	Phrenic Nerve Injury
QoL	Quality of Life
SAP	Statistical Analysis Plan
SF-12	Medical Outcomes Short Form-12
SoC	Standard of Care

3. Introduction

The purpose of the Statistical Analysis Plan (SAP) for the STOP Persistent AF study is to provide pre-analysis documentation and rationale for the statistical procedures used in planned analyses which are performed throughout the investigation. Specifically, this plan outlines methods employed in the study's Pre-Market Approval (PMA) submission and final report, as well as in any planned publications. 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 5 of the STOP Persistent AF Clinical Investigation Protocol (referred to as the CIP in this SAP), dated 17/APR/2017. 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. Study Objectives

4.1. Primary Efficacy Objective

Demonstrate an acceptable efficacy success rate at 12 months after the pulmonary vein isolation (PVI) ablation procedure.

4.2. Primary Safety Objective

Demonstrate an acceptable safety profile of the pulmonary vein isolation (PVI) ablation procedure.

4.3. Secondary Objective

Demonstrate an improvement in quality of life between baseline and 12 months after the index ablation procedure as measured by the AFEQT and SF-12 questionnaires.

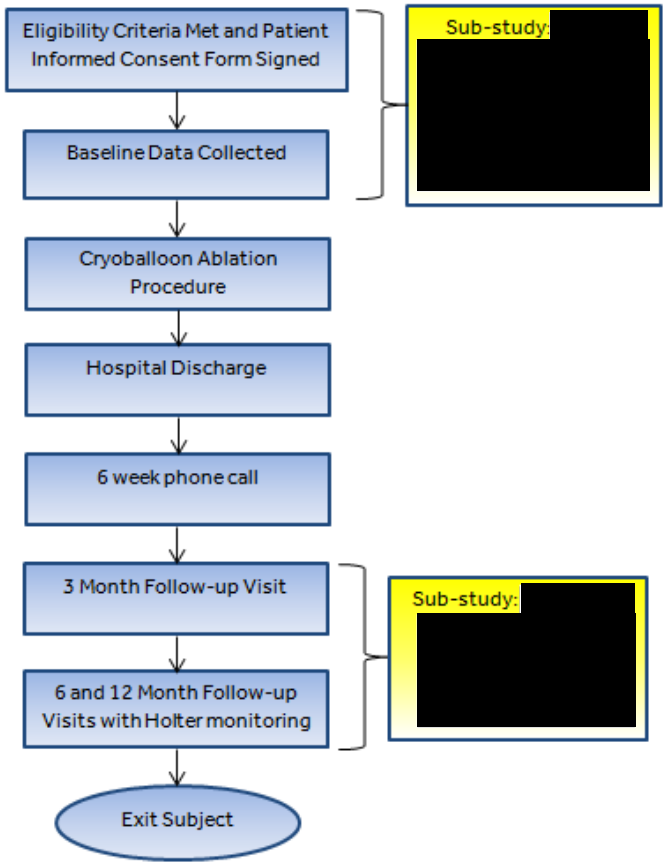
4.4. Ancillary Objectives

[redacted]

5. Investigation Plan

The purpose of the study is to provide data demonstrating the safety and effectiveness of the Arctic Front Advance and Freezor MAX Cardiac CryoAblation Catheters for the treatment of symptomatic drug refractory persistent AF. The study is proposed by the sponsor as a result of feedback received from the US Food and Drug Administration (FDA) on pre-submission Q151184, regarding a proposal to expand the indications for use for the Arctic Front Advance to include patients with persistent AF. The study is also designed to expand the indication for the Freezor MAX Cardiac CryoAblation Catheter. The proposed indication for the Arctic Front Advance CryoAblation Catheter is as follows: The Arctic Front Advance Cardiac CryoAblation Catheters are indicated for the treatment of drug refractory recurrent symptomatic paroxysmal and persistent atrial fibrillation. The proposed indication for the Freezor MAX Cardiac CryoAblation Catheter is as follows: The Freezor MAX Cardiac CryoAblation Catheters are indicated for use as an adjunctive device in the endocardial treatment of paroxysmal and persistent atrial fibrillation in conjunction with the Arctic Front Advance Cardiac CryoAblation Catheter for the following uses: gap cryoablation to complete electrical isolation of the pulmonary veins, cryoablation of focal trigger sites and creation of ablation line between the inferior vena cava and the tricuspid valve. A sub-study [redacted]

Medtronic, Inc. is sponsoring the STOP Persistent AF Study; a prospective, interventional, multi-center, non-randomized, single arm, unblinded clinical study. [redacted] Up to 225 subjects will be enrolled world-wide. In the US, and Canada, up to 200 subjects will be enrolled to ensure 150 subjects are treated with an Arctic Front Advance Cardiac CryoAblation Catheter. The maximum number of subjects treated at Canadian centers combined is 45 subjects. Up to 25 subjects will be enrolled in Japan to ensure 15 subjects are treated with an Arctic Front Advance Cardiac CryoAblation Catheter. The maximum number of subjects treated at Japanese centers is 15 subjects.



6. Statistical Methods

6.1. Study Subjects

6.1.1. Disposition of Subjects

Subjects included in our primary analysis cohort require all of the following conditions: eligibility, consent to be studied, and treatment pulmonary veins with the Arctic Front Advance Cardiac Cryoablation Catheter System. The study will collect data on subjects who consented to be studied and to the cryoablation procedure, but who were given a different treatment at the time of operation, but these cases will not affect the analysis of primary, secondary, or ancillary objectives, as they do not meet the criteria of undergoing a cryoablation procedure.

6.1.2. Clinical Investigation Plan (CIP) Deviations

Protocol deviations will be described using frequency tables and listings. At stated above in section 6.1.1, subjects will be excluded from analysis of primary, secondary, or ancillary endpoints if any of these conditions apply:

- Subject does not meet the study's entry criteria and exited prior to procedure
- Subject did not consent to the procedure
- Subject did not undergo a planned procedure with the Arctic FrontTM Cardiac Cryoablation Catheter System.

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 sign informed consent. The full analysis set will be used for AE reporting in general and ancillary objective [redacted]

The **Modified Intention-to-Treat** (mITT) cohort is the subset of subjects who maintain informed consent at least until the index cryoablation procedure was finished. For endpoints analyzed in this group of subjects, the standard Intention-to-Treat (ITT) protocol applies immediately upon insertion of the cryoballoon catheter into the subject's vasculature. The mITT cohort will be used in the analysis of the primary efficacy objective, the secondary objective regarding Quality of Life (QoL), and the ancillary objective [redacted]. The mITT dataset will be used in the analysis of the primary efficacy and safety objectives, as well as for the secondary objective and ancillary objectives[redacted].

This study has planned submissions to the US FDA and Japan PMDA. The Pre-Market Approval Submission (PMA-S) to the US FDA will include subjects from centers in US and Canada. If full 12-month data is available from Japan subjects at the time of PMA-S submission, the Japan data may be included in the PMA-S submission. The submission to the Japan PMDA will include analysis from data collected from all four geographies. The analyses for both submissions will follow the same pre-specified analysis procedures, so this SAP will not distinguish between the PMDA submission cohort (165 subjects expected) vs. the subset of subjects used for the PMA-S (150 subjects expected) except when differences due to sample size are relevant to the discussion.

6.2. General Methodology

6.2.1. Overview

This study follows patients with drug refractory symptomatic persistent AF who treated by cryoballoon ablation of the pulmonary veins for 12 months. The two co-primary endpoints of the study focus on the safety of the procedure and efficacy of the procedure in treating AF. The efficacy endpoint is subject to a three-month blanking period in accordance with the current FDA guidance[1].

[1] https://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm182016.htm#s6

The analysis described in this SAP will be conducted by Medtronic statisticians. Efficacy endpoints will be adjudicated by a core lab. Safety endpoints will be adjudicated by the Clinical Event Committee (CEC). Determination of events and event onset dates used in the PMA supplement, PMDA submission, and any resulting publications will be based on the core lab and CEC adjudications. CEC determinations will also be used to determine AE severity and procedure or device-relatedness. Prior to evaluation of the study's primary objectives, a descriptive analysis will be performed. Demographic and other key baseline characteristics will be summarized for the mITT dataset.

6.2.2. Primary Efficacy Objective

Demonstrate an acceptable efficacy success rate at 12 months after the pulmonary vein isolation (PVI) ablation procedure.

6.2.2.1. Hypothesis

The following hypothesis will be tested in a one-sided test at the 0.025 significance level:

Ho: $\pi_s \leq 40\%$

Ha: $\pi_s > 40\%$

where π_s is the probability of treatment success at 12 months.

6.2.2.2. Endpoint Definition

Treatment success is defined as freedom from treatment failure. Treatment failure is defined as any of the following components:

- Acute procedural failure
- Documented AF/AT/ AFL on Holter/TTM/12-lead ECG after the 90 day blanking period
  - Minimum of 30 seconds on Holter/TTM and 10 seconds on 12-lead ECG
- A reablation for the treatment of recurrent AF/AT/AFL after the 90 day blanking period
- Class I or III antiarrhythmic drug (AAD) dose increase from the historic maximum ineffective dose (prior to the ablation procedure) or initiation of a new Class I or III AAD after the 90 day blanking period. Note: remaining on the same pre-ablation dose or decreased dose, or re-initiation of a previously failed or not tolerated Class I or III AAD after the 90 day blanking is not considered a failure. Subjects are allowed to remain on Class I or III antiarrhythmic medications at the historic maximum ineffective dose (on prior to the ablation procedure) after the 90 day post-procedure blanking period.
- Ablation using RF in the left atrium

Blanking period is defined as the first 90 days after the index ablation procedure. 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 blanking period, recurrent arrhythmias can be managed with antiarrhythmic drugs, cardioversion or one cryo re-ablation procedure of the pulmonary veins. Titration of Class I and III antiarrhythmic medications are allowed during the blanking period.

Acute procedural failure is defined as:

- Inability to isolate all accessible targeted pulmonary veins (minimally assessed for entrance block and, where assessable, exit block) during the index procedure
- Left atrial non-PVI ablations including but not limited to, ablation of linear lesions, complex fractionated electrograms or non-PV triggers

6.2.2.3. Analysis Methods

The probability of a subject achieving effectiveness success at 12 months (365 days) will be estimated using survival analysis, the Kaplan-Meier method. The standard error will be approximated using Greenwood's formula. A two-sided 95% log-log confidence interval for the probability will be constructed. For every treated subject, day 0 is defined as the day of the index cryoablation procedure. 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 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 known study visit date.

For the component of the endpoint, documented AF/AT/AFL, if this documentation resulted from rhythm monitoring occurring at the 12-month visit within the 12-month visit window, the date of recurrence will be set to 365 days from the study ablation procedure so that these events will be counted as treatment failures in the 12-month Kaplan-Meier analysis. This objective will be analyzed in the mITT dataset: all enrolled subjects who have the Arctic Front Advance Cardiac CryoAblation Catheter inserted into vasculature will be included.

6.2.2.4. Performance Requirements

If the lower bound of the 95% confidence interval for the primary endpoint at 12 months is greater than the performance goal of 40%, the objective will be considered met. Rationale for the choice of performance goal is included in the CIP.

6.2.2.5. Sensitivity Analyses

A sensitivity analysis will be conducted to estimate the potential impact of subjects with less than 12 months of follow-up at the final analysis, which is described in more detail in section 6.5.

In the event a subject is enrolled, treated, but later discovered they did not meet eligibility, a sensitivity analysis will be conducted to understand the impact of these subjects. The same analysis methods will be utilized as described in section 6.2.2.3, but those subjects not meeting eligibility will be excluded.

6.2.3. Primary Safety Objective

Demonstrate an acceptable safety profile of the pulmonary vein isolation (PVI) ablation procedure.

6.2.3.1. Hypothesis

The following hypothesis will be tested in a one-sided test at the 0.025 significance level:

Ho: PS ≥ 13%

Ha: PS < 13%

Where PS is the probability of a safety event through 12 months.

6.2.3.2. Endpoint Definition

A primary safety event is defined as a serious procedure-related or serious system-related adverse event including the following:

- Transient ischemic attack (within 7 days of ablation procedure)
- Cerebrovascular accident (within 7 days of ablation procedure)
- Major bleeding that requires transfusion (within 7 days of ablation procedure)
- Cardiac perforation, tamponade or pericardial effusion (within 7 days of ablation procedure)
- Pulmonary vein stenosis (>75% reduction within 12-months of ablation procedure)
- Myocardial infarction (within 7 days of ablation procedure)
- Phrenic nerve injury (unresolved at 12-months)
- Atrio-esophageal fistula (within 12-months of ablation procedure)
- Death (within 7 days of ablation procedure)

6.2.3.3. Analysis Methods

The probability of a safety event at 12 months (365 days) will be estimated using survival analysis, the Kaplan-Meier method. The standard error will be approximated using Greenwood's formula. A two-sided 95% log-log confidence interval for the probability will be constructed.

For every treated subject, 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, those subjects will be censored at the last study contact date recorded on CRF which may include the last study visit, the exit date, or death date. If a subject without a safety event is lost to follow-up, the censoring date will be set to the last known study visit date.

For subjects with a repeat ablation within 12 months, the start of the survival analysis will not reset. Day 0 will remain the day of the index cryoablation procedure. Safety events related to the repeat ablation procedure occurring on or prior to 365 days post the index cryoablation procedure will be counted as safety events and count against the primary safety objective. This objective will be analyzed in the mITT dataset: all enrolled subjects who have the Arctic Front Advance Cardiac CryoAblation Catheter inserted into vasculature will be included.

6.2.3.4. Performance Requirements

If the upper bound of the two-sided 95% confidence interval at 12 months is less than the performance goal of 13%, the objective will be considered met.

6.2.4. Secondary Objective

Demonstrate an improvement in quality of life between baseline and 12 months as measured by the AFEQT (Atrial Fibrillation Effect on QualiTy-of-life) and Medical Outcome Short Form-12 (SF-12) questionnaires.

The secondary objective, Quality of Life, will be evaluated to gain additional information about the performance of the Arctic Front Advance Cardiac CryoAblation catheter. There are three hypotheses

tested in the objective (separate hypothesis tests related to the AFEQT questionnaire, the physical component score of the SF-12 questionnaire, and the mental component score of the SF-12 questionnaire). A Hommel multiple testing procedure will be utilized to maintain an overall type I error rate of 0.025 for this objective2. The simultaneous analysis of secondary objectives is described in more detail in section 6.5.

6.2.4.1. AFEQT

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 corresponds to complete disability and 100 corresponds to no disability.

Hypothesis

The following hypothesis will be tested in a one-sided test at the 0.025 significance level:

Ho: ΔAFEQT= 0

Ha: ΔAFEQT > 0

Where ΔAFEQT is the change in AFEQT score from baseline to 12 months.

Analysis Methods

Change in AFEQT score is defined as 12-month AFEQT score minus baseline AFEQT score. Change in AFEQT scores will be assessed utilizing a one-sample t-test. A two-sided 95% confidence interval will be calculated based on the t-distribution.

Additionally, summary statistics (e.g. mean, SD, median, range) and graphical methods will be used to summarize the change in AFEQT scores from baseline through 12 months. The outcome variable for this objective requires both a baseline and a follow-up assessment of AFEQT scores with answers for all questions. Analysis will be performed on the mITT dataset for those subjects that have completed at least 50% of the questions for each subscale at baseline and 12-months.

Performance criteria

If the p-value from the one-sample t-test after adjusting for the Hommel procedure is < 0.025, the objective will be considered met.

Additional Analyses

The AFEQT questionnaire has three subscale scores, 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 based on the t-distribution.

[2] Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988; 75, 383-386.

In addition to assessments of AFEQT scores at baseline and 12-month follow-ups, subjects will also complete an AFEQT survey at their 6-month visit. AFEQT data from the six-month visit will be included in graphical displays and descriptive summaries relating to this objective, but will not be included in the hypothesis test for this endpoint, which is change from baseline to 12 months. However, in addition to the test about the primary endpoint, we may perform and report on an exploratory repeated measures analysis of composite scores (e.g., using a linear mixed effects model with a subject-level random effect and unstructured covariance) to assess the change in QoL between baseline and six months, and between six months and 12 months. This may be accomplished using SAS proc mixed code such as:

6.2.4.2. Medical Outcome Short Form-12 (SF-12)

The Medical Outcome Study Short Form-12 (SF-12) questionnaire will also be utilized for this objective. The SF-12 questionnaire is a health-related quality of life questionnaire to evaluate the subject's mental and physical performance. Physical and mental health component scores are calculated using responses to 12 questions with a response range from 0 – 100, with 0 corresponds to lowest level of health and 100 indicates highest level of health.

Hypothesis

The following hypotheses will be tested. Each will be tested in a one-sided test at the 0.025 significance level:

Ho: ΔSF-12mental = 0 Ho: ΔSF-12physical= 0
Ha: ΔSF-12mental > 0 Ha: ΔSF-12physical > 0

Where ΔSF-12mental is the change in ΔSF-12 mental score from baseline to 12 months, and ΔSF-12physical is the change in ΔSF-12 physical score from baseline to 12 months

Analysis Methods

Change in SF-12 component score is defined as 12-month SF-12 score minus baseline SF-12 score. Change in SF-12 scores will be assessed utilizing a one-sample t-test. A two-sided 95% confidence interval will be calculated based on the t-distribution.
Additionally, summary statistics (e.g. mean, SD, median, range) and graphical methods will be used to .summarize the change in SF-12 scores from baseline through 12 months. The outcome variable for this hypothesis test requires both a baseline and a follow-up SF-12 form to be completed. For a subject's physical health score to be calculated at either time point, all physical health dimension questions must be answered; likewise, for mental health scores to be calculated, all mental health questions must be

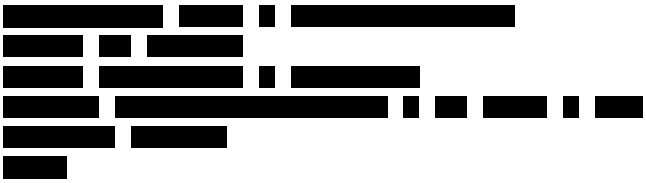
answered. Analysis will be performed on complete cases within the mITT dataset, in subjects that have completed both a baseline and 12-month questionnaire.

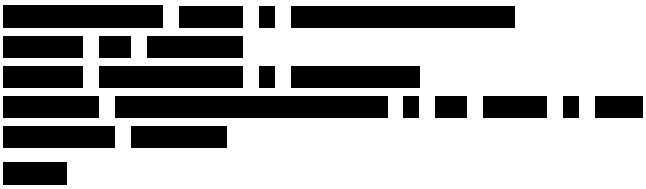
Performance criteria

If the p-value from the one-sample t-test after adjusting for the Hommel procedure is < 0.025, the objective will be considered met.

Additional Analyses

As with AFEQT assessments, SF-12 scores will be assessed as 6-month visits in addition to baseline and 12-month visits. While only the baseline and 12-month assessment will be used in the hypothesis tests for one-year change in SF-12 physical and mental health scores, all graphical and descriptive presentations of SF-12 outcomes will show data from all three timepoints. Additionally, results from a repeated measures analysis using all available data may be shown in the final report, to provide information on the degree of change in QoL in the first and second half years after the cryoablation procedure. The following code for SAS proc mixed provides a template for this analysis:





6.2.4.3. Sensitivity Analyses

Based on FDA correspondence related to written feedback for Arctic Front family of cryoballoon catheters; Freezor Max cardiac cryoablation catheter (Q181709/S001; received by Medtronic January 18, 2019), sensitivity analyses will be performed to assess the possible effects of missing AFEQT and SF-12 scores upon the hypotheses that are tested. For both outcomes, sensitivity analyses will investigate a "worst-case" scenario, in which missing responses for questions at baseline are assumed to be from the highest possible level (leading to the highest possible group average score at baseline). Missing responses to survey questions at 6-month and 12-month visits are assumed to be values indicating the lowest levels of mental and physical QoL. Hypothesis testing will proceed for the AFEQT using the same methods outlined in section 6.2.4.1, and will use the methods outlined in section 6.2.4.2 for the SF-12. If the statistical significance does not differ between the worst-case and complete case analyses, it will be concluded that the findings are robust to missing data. Meanwhile, if the worst-case scenario from the observed complete-cases scenario differs in terms of statistical significance, the sensitivity will be noted, and further investigation may occur while utilizing less extreme assumptions about missing data patterns.

6.2.5. Ancillary Objectives

Ancillary objectives been defined to provide additional information about the performance of the Arctic Front Advance Cardiac CryoAblation Catheter. For these objectives, no hypotheses are defined for regulatory or labeling purposes.

[redacted]

[redacted]

[redacted]

[redacted]

6.3. Center Pooling

The STOP Persistent AF study is expected to be conducted at up to 25 study centers in the US, Canada, Europe and Japan, with a total enrollment of 225 subjects across all study centers to ensure that at least 150 are treated with the Arctic Front Advance Cardiac CryoAblation Catheter. Per the CIP, each site will treat between 0 and 15 subjects. The protocol also specifies that a combined total of no more than 45 subjects will be enrolled at Canadian Centers. The target total enrollment for all Japanese centers combined is 15 subjects.

Evidence for this study will be used in submissions to two separate regulatory bodies. The submission for indication change to the Japan Pharmaceuticals and Medical Devices Agency (PMDA) will include analysis based on subjects from all available geographies (US, Canada, and Japan), and is referred to as the PMDA cohort in this document. Meanwhile, the Pre-Market Approval Supplement (PMA-S) to the US Food and Drug Administration (FDA) will include data from subjects treated in the US, and Canada; these subjects are referred to as the "PMA-S" cohort in this SAP. For both PMA-S and PMDA cohorts, it will be assessed whether site-to-site and geographical heterogeneity exists in the rate of pre-specified primary endpoints using a random-effects meta-analytic approach. The R statistical software platform will be used to evaluate whether sites exhibit significant heterogeneity in event rates, and whether geography (a three-level categorical variable representing whether a site is located in the US, Canada, or Japan) moderates any statistically significant heterogeneity that is observed. Models will be fit separately for each primary outcome. Testing will not be performed for outcomes or strata in which heterogeneity cannot be assessed in a meaningful way (for example, if < 5 events are observed in all subjects). If a Cochran's Q-test for heterogeneity shows $p<0.05$, it will be taken as evidence of significant heterogeneity between sites. Evidence of between-site heterogeneity will not preclude pooling data; rather, it will prompt further investigation into the sources of the apparent differences in event rates between sites. At a minimum, findings of analyses on heterogeneity between study sites and between all US, Canadian and Japanese study centers will be shown in the final report in a table by endpoint. Heterogeneity analysis findings will not change the composition of the PMA-S and PMDA cohorts as defined above. Analysis of heterogeneity will be performed within the mITT dataset, like the primary safety and efficacy analyses are.

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

The mortality rate in this patient population is anticipated to be low. The total assumed attrition rate through one year is 10%. 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 with less than 12-months of 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, tipping point analyses will be performed on the full analysis set.

The tipping point analysis methods will be as follows. For subjects in the mITT dataset with less than 12-months of follow-up, each subject will be set to treatment failure (failure date set to date of study exit + 1 day). The 12-month Kaplan-Meier estimate will be re-calculated with the earliest (closest to index ablation) sequentially added to the Kaplan-Meier analysis, and the results presented in table format. The tipping point will be defined as the number of additional failures in which the 40% OPC is crossed by the lower 95% confidence bound.

Those subjects with less than 12 months of data will be compared with completers in terms of baseline covariates, and results will be displayed in table form. 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 Hommel multiple testing procedure will be utilized to maintain an overall type I error rate of 0.025 for the three hypotheses tested as part of the secondary objective3. Testing for this objective will be performed if the primary objectives are met. The Hommel procedure is a "step-up" method of error adjustment.

The three hypotheses associated with the secondary objective (see 6.2.3) will be denoted here as H(1), H(2), and H(3). H(1), H(2) and H(3) denote the null hypothesis for change in AFEQT, change in SF-12 mental component, and SF-12 physical component. For each of the hypotheses, p-values will be calculated and sorted $p(1) < p(2) < p(3)$. The decision rule to accept or reject each hypothesis will follow the step-up algorithm, where $\alpha=0.025$.

Step 1: If $p(3) > \alpha$, accept H(3) and go to Step 2, otherwise reject all hypotheses and stop

Step 2: If $p(2) > \alpha/2$, accept H(2) and go to Step 3, otherwise reject all remaining hypotheses and stop

Step 3: If (1) $\alpha/2 < p(2) \leq 2\alpha/3$ and $p(1) \leq \alpha/2$ or (2) $p(1) \leq \alpha/3$, reject H(1); otherwise accept H(1).

[3] Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988; 75, 383-386.

Assume that AFEQT scores, SF-12 mental health scores, and SF-12 physical health scores are respectively abbreviated as AFEQT, SF12ment and SF12phys, and that an indicator 'FollowUp' denotes whether the three QoL outcome variables was taken at a baseline or annual follow-up visit. The above procedure can be implemented in SAS Proc Multtest to produce multiplicity-adjusted p-values with code such as this:

[redacted]

If this resulted in adjusted p-values of < 0.025, the associated outcomes would be considered met.

6.6. Interim Analyses

No interim analyses are planned. At the completion of the last 12 month visit from the enrolled and treated US and Canada subjects, a PMA-S will be submitted to the FDA. Study data from the Japanese centers may not be included in the PMA-S supplement. The submission for the Japan PMDA will occur after the PMA-S without regard to the findings in the PMA-S. In both cases, analysis will occur only at 100% of the total information fraction for the cohort, eliminating the necessity of a type I error spending control technique for multiple analyses of the data.

6.7. Subgroup Analyses

A limited number of additional analyses will be performed to evaluate evidence for a differential effect of cryoablationon the primary and secondary endpoints within subgroups of subjects.
At the time of this SAP, current FDA guidance recommends additional evaluation of primary objectives and secondary objective within the following demographic subgroups:

- Age (calculated as [year of baseline 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. The following code for SAS Proc Lifetest can be used as a general template for subgroup analysis:

[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 information sparsity. Age will be calculated as [year of baseline 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. The log-rank test will be utilized. 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 or secondary objectives are analyzed.

In addition to subgroup analyses by demographic variables, there are eight pre-specified subgroup analyses that will be performed:

- by Left Atrial Diameter (LAD) at the most recent measurement prior to the procedure;
- by duration of persistent AF, measured in terms of months prior to the procedure;
- by number of prior failed AAD's
- by history of hypertension (binary)
- by history of cardioversion (binary)
- by Subjects with baseline ECG performed within 30 days of index ablation versus > 30 days.
- by body mass index (BMI) > 40 versus ≤ 40
- Major inclusion criteria: Subjects with symptomatic persistent AF documented by two documented ECGs versus subjects with symptomatic persistent AF documented by one ECG and a doctor note indicating the patient had symptoms consistent with AF

The last three subgroup analyses in the list above (by ECG within 30 days of baseline, by BMI > 40, and by major inclusion criteria documentation method) were added based on FDA's correspondence related to STOP Persistent AF IDE Study (G160177/S007) received by Medtronic on January 10, 2018. Each of these subgroups will be analyzed using nonparametric Kaplan-Meier methods like the demographic variables in the manner described above. LAD will be divided into quartiles, while divisions of prior persistent AF duration and prior failed AAD's into discrete bands will be determined after examination of the distribution of these variables in the data.

6.8. Changes to Planned Analysis

Comprehensive details on analysis methods have been included to this SAP, but no changes to the statistical methods or principal features defined in the STOP AF First CIP version 5 are noted. However, there are two minor additions from the methods defined in the STOP Persistent AF CIP version 5. First, clarification has been added that while this study assesses subjects' QoL at six month follow-ups, and results from 6-month assessments will be made available in graphs and tables, the endpoints for the secondary objective are defined based only on QoL at baseline and 12 months. An additional repeated measures analysis may be performed outside of the analysis pre-specified for the secondary objective that utilizes data from all three points of assessment. Second, [redacted] was added to ancillary objective [redacted]

[redacted] Additional analysis or deviations from procedures in this SAP may be addressed by the release of newer SAP versions, 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 Persistent AF was determined by finding the minimum number of subjects enrolled which provides adequate power 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 Efficacy Endpoint

Sample size was estimated using Proc Power in SAS 9.4. The calculation was designed to find the number of analyzed subjects required for a one-sided exact binomial test (α= 0.025) to have 90% power in declaring the event rate significantly greater than the efficacy performance goal of 40%, when in fact the true rate of treatment success at 12 months is 54%. Attrition from enrollment to 12-month follow-up was assumed to be 10%, implying 150*(1 – 0.1) = 135 and floor(165*(1 – 0.1)) = 148 subjects analyzed at one year for the PMA-S and PMDA datasets, respectively. The sample size calculation under these conditions can be replicated with the following SAS code:

[redacted]

Computed Power

	N Total	Upper Crit Val	Actual Alpha	Power
(PMA-S)	135	66	0.0224	0.899
(PMDA)	148	72	0.0202	0.917

The power calculation outlined above is based on the binomial test, which is conservative compared to the power provided by the Kaplan-Meier methods (described in 6.2.2) that will actually be used. In the case of Kaplan-Meier analysis, 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) will provide no information to the analysis.

7.1.1. Rationale for Performance Criteria

Full details of the rationale for performance criteria can be found in the CIP. The choice of 12-month follow-up and acceptable success rate of 40% performance criteria was selected based on the 2012 HRS/EHRA/ECAS Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation: Recommendations for Patient Selection, Procedural Techniques, Patient Management and Follow-up, Definitions, Endpoints, and Research Trial Design, provides recommendations for success rates in clinical trials. The recommendation for evaluating the efficacy of a treatment for persistent AF is as follows: "*If minimum chronic success rate is selected as an objective effectiveness endpoint for a clinical trial, we recommend that the minimum chronic acceptable success rate for persistent AF at 12-month follow-up is 40%.*"

7.2. Sample Size Determination for Safety Endpoint

Sample size for the primary safety objective was estimated using Proc Power in SAS 9.4. Parameters for the safety endpoint power calculation were a one-sided exact binomial test with α = 0.025 to declare a significant difference between a performance goal (PG) of 13% when the true safety event rate was 5.0% and the sample size was 135 – 148 subjects analyzed. The following code for proc power can be used to replicate the sample size calculation under these conditions.

[redacted]

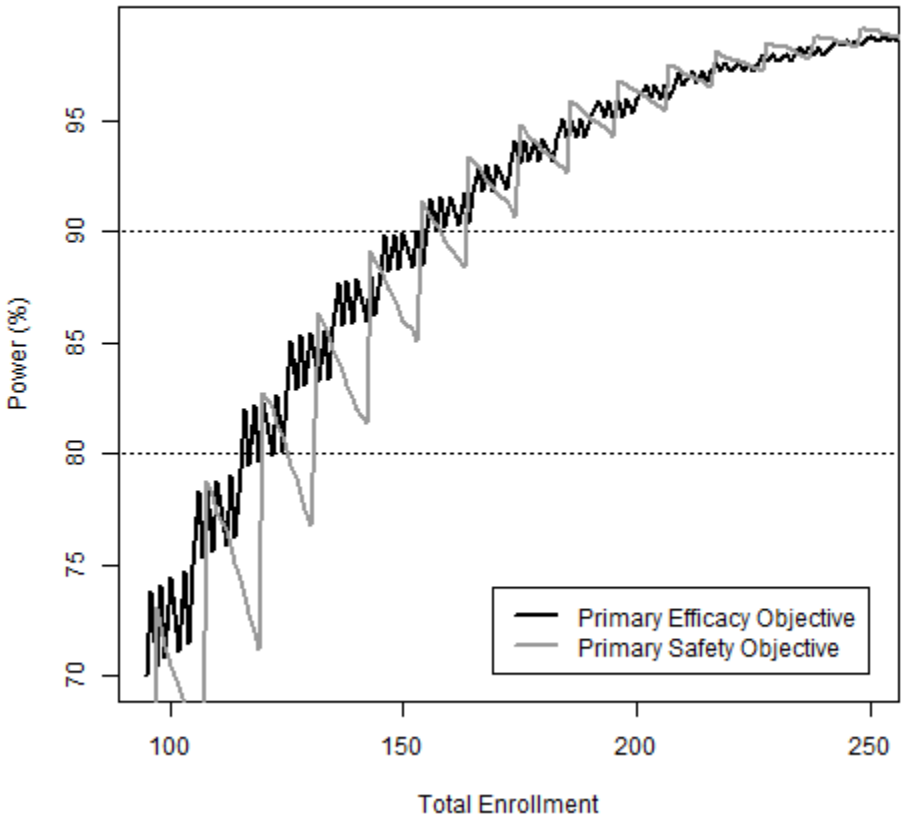
Computed Power

	N Total	Lower Crit Val	Actual Alpha	Power
(PMA-S)	135	9	0.0141	0.860
(PMDA)	148	11	0.0233	0.932

This calculation shows that with 10% attrition between enrollment and 12-month follow-up visits, 150 subjects in the US and Canada will provide 86% power to accept the alternative hypothesis for a rate of 5%. The power for the PMDA dataset, with an estimated 15 extra subjects (for 165 total), will be 93% for the primary safety objective.

Power for both endpoints by sample size (accounting for attrition, and assuming the constraints mentioned above) is shown below in figure 2:

Figure 2:
Power vs. Enrollment for STOP Persistent AF Primary Endpoints



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 and ancillary objectives will be validated with a minimum of level II validation. Analyses that are not related to primary objectives or ancillary endpoints will be validated at a minimum of level II validation if being presented externally in an abstract or publication.

9. References

Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 1988; 75, 383-386.

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. "Clinical Study Designs for Surgical Ablation Devices for Treatment of Atrial Fibrillation – Guidance for Industry and Food and Drug Administration Staff". FDA Web Site. https://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm182016.htm#s6. Published February 15, 2013. Accessed May 26, 2017.