# **C**atheter **Ab**lation Versus **An**tiarrhythmic Drug Therapy for **A**trial Fibrillation Trial – Statistical Analysis Plan

NCT: 00911508

Date: 26 February, 2018

# Statistical Analysis Plan for CABANA

Analysis Plan dated February 26, 2018
Duke Clinical Research Institute

**Protocol Date:** November 22, 2013 (Version 3.5)
**Sponsor:** National Heart, Lung, and Blood Institute

---

# Table of Contents


1. Study Title ..... 3
2. Study Overview ..... 3
3. Objectives ..... 3
   3.1 Primary Objective ..... 3
   3.2 Secondary Objectives ..... 4
4. Subject Randomization ..... 4
5. Primary Analysis Population ..... 4
6. Data Sources ..... 5
7. General Analysis Methodology ..... 5
   7.1 Demographic and Baseline Analyses ..... 6
   7.2 Disposition of Subjects ..... 6
   7.3 Missing Data ..... 6
8. Primary Endpoint Analysis ..... 7
   8.1 Other Analyses of Primary Endpoint Data (as randomized): ..... 8
9. Analysis of Secondary Endpoints ..... 9
   9.1 All-cause mortality (key secondary endpoint) ..... 9
   9.2 All-cause Mortality or Cardiovascular Hospitalization ..... 10
   9.3 All-cause Mortality, Stroke, or Hospitalization for Heart Failure or Acute Ischemic Events ..... 11
   9.4 Cardiovascular Death ..... 11
   9.5 Cardiovascular Death or Disabling Stroke ..... 12
   9.6 Arrhythmic Death or Cardiac Arrest ..... 13
   9.7 Heart Failure Death ..... 13
   9.8 Recurrent Atrial Fibrillation ..... 14
   9.9 Cardiovascular Hospitalization ..... 15
   9.10 Medical Costs, Resource Use and Cost Effectiveness. ..... 17
   9.11 Quality of Life. ..... 17
   9.12 Major Non-Endpoint Adverse Events. ..... 17
   9.13 LA Size, Morphology and Function. ..... 17
10. “On-Treatment” Analysis ..... 18
11. Per-Protocol Analysis ..... 19
12. Multiple Comparisons ..... 19
13. Interim Analyses ..... 20
References ..... 22

# 1. Study Title

**C**atheter **Ab**lation Versus **An**tiarrhythmic Drug Therapy for **A**trial Fibrillation (**CABANA**)

# 2. Study Overview

CABANA is a multi-center randomized clinical trial to assess the safety and efficacy of percutaneous left atrial catheter ablation versus antiarrhythmic drug therapy in subjects who are at least 18 years of age and have new onset or under-treated paroxysmal, persistent, or long-standing persistent atrial fibrillation (AF) and warrant therapy for their arrhythmia. The study population consists of 2,204 subjects enrolled at 126 clinical sites over a period of approximately six years. Subjects were randomized in equal proportions (1:1) to receive either catheter ablation or drug therapy and followed at regular intervals for the duration of the study. The minimum length of follow-up will be slightly less than 2 years, and the median duration of follow-up will be approximately 4 years. At study entry and during the follow-up period, extensive clinical, quality-of-life, and economic data are collected.

# 3. Objectives

## 3.1 Primary Objective

The primary objective of CABANA is to determine whether the treatment strategy of percutaneous left atrial catheter ablation for the purpose of eliminating atrial fibrillation will improve clinical outcomes compared to treating the arrhythmia with current state-of-the-art drug therapy consisting of either rate control or rhythm control drugs.

The **primary endpoint** is the time to the first event among a composite of major clinical outcomes consisting of

- Death (all-cause)
- Disabling stroke
- Serious bleeding
- Cardiac arrest

The primary hypothesis of the study is that the treatment strategy of percutaneous left atrial catheter ablation will reduce the incidence of this composite endpoint in patients with untreated or under-treated AF warranting therapy, compared to current state-of-the-art drug therapy.

### 3.2 Secondary Objectives

The secondary objectives are to compare the following clinical, economic, and quality-of-life outcomes in subjects randomized to catheter ablation versus state-of-the-art drug therapy:

1. **All-cause mortality (key secondary endpoint)**
2. All-cause mortality or cardiovascular hospitalization
3. All-cause mortality, stroke, or hospitalization for heart failure or for an acute ischemic event
4. Cardiovascular death
5. Cardiovascular death or disabling stroke
6. Arrhythmic death or cardiac arrest
7. Heart failure death
8. Recurrent atrial fibrillation
9. Cardiovascular hospitalization
10. Medical costs, resource utilization, and cost effectiveness
11. Health-related quality-of-life
12. Composite adverse events
13. Left atrial size, morphology, and function

To aid in the interpretation of the primary and several of the secondary endpoint comparisons, the individual components of the primary endpoint (in addition to death) will also be summarized between the randomized arms of the trial.

## 4. Subject Randomization

The treatment modalities being studied in CABANA are clinically-established treatment strategies that are used routinely across the U.S. and around the world. No experimental drug or treatment is involved. The trial intervention is simply the random assignment of the subjects to one or the other of the two treatment strategies.

Eligible subjects who provided written informed consent and met all inclusion and no exclusion criteria were randomly assigned in equal proportions (1:1) to either the catheter ablation arm or the drug arm using permuted blocks, stratified by clinical site. Subject randomization was accomplished by telephone through a centralized toll-free Interactive Voice Response System (IVRS).

## 5. Primary Analysis Population

Except where otherwise noted, all randomized subjects will be included in the analysis population for assessing the primary and secondary endpoints. All major treatment comparisons between the randomized groups in the trial will be performed according to the principle of "intention-to-treat" (ITT); that is, subjects will be analyzed and endpoints attributed according to the treatment arm to which the patients were randomized, regardless of subsequent crossover or

post-randomization medical care. As explained later in this document (section 10), some additional analyses will be conducted using "on-treatment" data.

## 6. Data Sources

The enrolling clinical sites complete a series of electronic Case Report Forms (eCRF) for each subject randomized, including demographic, clinical, and quality of life forms at enrollment and at specified follow-up intervals (the follow-up intervals are detailed in the study protocol), event forms for death, neurological events, bleeding, cardiac arrest, and other ancillary forms as required.

Clinical data are collected using the InForm electronic data capture (EDC) system. Except as noted below in section 9.8, rhythm data are collected using the "CABANA Box," a rhythm monitoring system provided for each patient by Medicomp. The rhythm monitoring data are sent from Medicomp to the Duke Clinical Research Institute (DCRI), where a determination is made as to which rhythms should be sent to the ECG core lab in Seattle for review and adjudication. The core lab sends adjudicated results back to the DCRI. Sites with any subjects not using the "CABANA Box" submit rhythm data on those subjects through a secure portal using the CABANA sharepoint website, which is then processed and submitted to the ECG core lab for adjudication. CT/MRI data are received and processed by Mayo Clinic's Imaging Analysis Laboratory, and the information is then transferred to the DCRI.

All data collected in the study are managed and analyzed at the CABANA Statistical and Data Coordinating Center at the DCRI. The InForm eCRF data are carefully reviewed, queried as needed, and then downloaded as raw SAS data files, and further review and checking of the data occur. The raw SAS data, analysis datasets, and analysis programs are stored on the secure DCRI statistical server. Final analyses will be performed at the DCRI using SAS version 9.4 or higher (SAS Institute Inc., Cary, NC). All programs written to create analysis datasets and perform analyses will be validated according to SOPs established by the DCRI Statistical Programming group.

## 7. General Analysis Methodology

As stated previously and except where otherwise noted, all major treatment comparisons will be performed according to the principle of "intention-to-treat." Subjects will be analyzed and endpoints attributed according to the treatment arm to which patients were randomized, regardless of subsequent crossover or post-randomization medical care.

Appropriate statistical models and tests will be used to examine the effect of the treatment strategies on the primary and secondary endpoints. For time-to-first-event endpoints, the log-rank test[1] and Cox proportional hazards model[2] will be used for statistically comparing the treatment groups unless otherwise specified. Additional perspective regarding the interpretation of the data will be provided through extensive use of confidence intervals and graphical displays, including plots of Kaplan-Meier[3] event-rate estimates. All statistical comparisons will be

performed using two-sided significance tests and a nominal level of significance of $\alpha$=0.05 unless otherwise specified.

### 7.1. Demographic and Baseline Analyses

Descriptive summaries of the distribution of continuous baseline variables will be presented in terms of percentiles (median, 25$^{th}$ and 75$^{th}$ percentiles), while discrete variables will be summarized in terms of frequencies and percentages. Statistical comparisons of treatment groups with respect to baseline characteristics will be limited to selected variables and disease factors known to influence prognosis. These variables will include: age, sex, race, descriptors of comorbidity, heart failure class, type of AF (paroxysmal, persistent or permanent), years since onset of AF, days from most recent qualifying episode of atrial fibrillation to enrollment, presence/absence of structural heart disease, and $CHA_2DS_2$-VASc score. For comparisons of treatment groups with respect to continuous or ordinal baseline variables, the nonparametric Wilcoxon rank-sum test will be used. Group comparisons with respect to discrete baseline variables will use the Pearson's chi-square test or Fisher's exact test as appropriate.

### 7.2 Disposition of Subjects

The disposition of subjects (number randomized, number who received the randomly assigned treatment, number of treatment crossovers, number of patients who withdrew consent, and number lost to follow-up) will be summarized overall and by treatment arm. The summary of treatment crossovers will distinguish the patients randomized to the drug arm who subsequently underwent catheter ablation (crossed over to ablation) during the course of their follow-up and the patients randomized to the ablation arm who never received an ablation. The distribution of the length of follow-up for the overall study population and for each randomized arm will be characterized in terms of percentiles (median, 25$^{th}$, and 75$^{th}$ percentiles). The number and percentage of patients who had at least 3 years of follow-up, and at least 5 years of follow-up will also be tabulated and reported. Among patients in the drug arm who crossed over to ablation, the distribution of length of follow-up prior to the crossover will be described in terms of percentiles. Also the distribution of follow-up among patients who withdrew consent (i.e., patients who withdrew prior to the protocol-specified 3 years of follow-up) will be summarized in terms of percentiles. Additionally, the number and percentage of patients who withdrew consent before 3 years, after 3 to <5 years, and after 5 years will be tabulated.

### 7.3 Missing Data

Missing data will be handled differently for covariates and endpoints. Data for covariates (see Section 8.1) are expected to be complete or near complete for all patients. Where missing data do occur, using single-imputation methods, continuous variables will be imputed using the median value and categorical variables will be imputed using the mode.

Most primary and secondary endpoints are time-to-event analyses which rely on the ability to calculate the time from randomization (or in some cases time from the end of the 90-day blanking period) to when the event occurred or the date of last contact for censored individuals. Here, missing data may occur when: 1) a visit form or question that indicates if an event occurred (or not) is not completed, and 2) a form or question indicates that an event occurred, but

the event date is missing. The first category prevents the assignment of whether an event occurred or not, while the second category prevents calculation of time to the event. In all analyses, we consider only non-missing data in the determination of whether or not an event occurred (i.e., cases in category 1 will not be considered). Time-to-event for known events when the event date is missing (category 2), will be assigned based on the visit with which the event is recorded. For example, if the event is reported on the month 3 visit, then the time to event would be set to 91.26 days (365/12 * 3). If the event is reported on the month 6 visit, then the time to event would be set to 182.52 days (365/12 * 6), and so on for all visits up to month 96. If patients do not experience an event, and have an event-free status associated with a visit but no visit date on their last known visit, they will be assigned a censor date in the same manner. For patients with multiple events including a death, any imputed non-death event dates (e.g., cardiovascular (CV) hospitalization) will be forced to occur before the date of death.

If an event or censored status is reported, but both the visit number and event/censored date are missing, then the date of event/censor will be imputed using relevant data and considered on a case-by-case basis.

## 8. Primary Endpoint Analysis

***Primary Endpoint:*** Time from randomization to the first event among a composite of major clinical outcomes consisting of

- Death (all-cause)
- Disabling stroke
- Serious bleeding
- Cardiac arrest

All possible events are carefully reviewed and adjudicated by a Clinical Events Committee (CEC) completely blinded to the randomized treatment assignment.

***Endpoint Definitions:*** Detailed definitions for each of the events are outlined in the CEC charter. We note that a disabling stroke during follow-up must be documented by a neurologic evaluation or brain CT / MR studies and result in a disabling, irreversible physical limitation defined by a Rankin Stroke Scale $\geq 2$. Serious (or life-threatening) bleeding is considered to have occurred if intracranial bleeding or other bleeding occurs at any time during follow-up accompanied by hemodynamic compromise that requires surgical intervention or a transfusion of $\geq 3$ units of blood. The serious bleeding definition *does not* include bleeding occurring with cardiac perforation or tamponade.

The time from randomization to the first event among the components of the primary composite endpoint will be measured (in days) for those who experienced an event and calculated as the date of the first event minus the date of randomization. For patients who do not experience any of the component events or who withdraw consent or drop out of the study before experiencing an event, time from randomization to the date of last contact will be used in the analysis, and those patients will be considered as censored observations in the time-to-event analysis. For

interim analyses, the best available endpoint data will be used (i.e., adjudicated data when available; otherwise site-reported event data).

***Statistical Methods:*** The log-rank test[1] will be the primary analytic tool for statistically assessing outcome differences between the two randomized treatment strategies with respect to the primary composite endpoint. This analysis will be conducted using PROC LIFETEST in SAS. A hazard ratio (HR) and 95% confidence interval (CI) summarizing the difference in outcome between the two randomized arms will be computed using the Cox proportional hazards model[2] (PROC PHREG in SAS), using treatment as the only predictor in the model.

In addition to the statistical hypothesis testing, cumulative event rates will be calculated according to the method of Kaplan and Meier[3] for each randomized arm as a function of time from randomization, and the estimated event probabilities will be displayed graphically.

The methods described above will constitute the primary analysis of this endpoint. The primary hypothesis is defined as:

$$H_0: \quad S_a(t) = S_d(t), \text{ vs.}$$

$$H_a: \quad S_a(t) \neq S_d(t),$$

where $S_a(t)$ is the event-free survivor function for the ablation arm of the trial, and $S_d(t)$ is the event-free survivor function for the drug arm, both of which are a function of follow-up time t. In other words, we are testing whether the event-free survival experience of the two arms of the trial is the same vs. whether it is different. The above hypothesis will be tested at a 2-sided significance level of 0.049 in the ITT population to account for a single interim analysis (see Section 13).

Because the Cox model will be used for calculating a hazard ratio and confidence interval to summarize the difference in outcome between the two randomized arms, the proportionality assumption in the Cox model (i.e., constant hazard ratio over time) will be checked and tested through adding an interaction term (treatment*log(t)) to the Cox model, and by examining a plot of Schoenfeld residuals.

The hypothesis as expressed above will also be used for the time-to-event analyses of secondary endpoints described below using a 2-sided significance level of 0.05.

### 8.1 Other Analyses of Primary Endpoint Data (as randomized):

***Treatment comparison with covariate adjustment.*** Using the Cox model, the effect of treatment strategy will be tested, adjusting for the following prospectively defined set of baseline patient characteristics that are expected to have a prognostic relationship with the primary endpoint: age (as a continuous variable), sex, minority status, baseline AF type, years since onset of AF, coronary artery disease, history of congestive heart failure, structural heart disease, hypertension, and $CHA_2DS_2$-VASc score. These characteristics will serve as a prelude to analyses for examining differential treatment effects. Structural heart disease will be defined as: history of valve disease with either surgical correction or percutaneous intervention, history of

cardiomyopathy, mitral valve regurgitation of moderate or greater severity, or mitral valve stenosis of moderate or greater severity.

***Differential treatment effects.*** If there is an overall difference in outcome between treatment groups, supplementary analyses will be performed to assess whether the therapeutic effect is consistent across all patients, or whether it varies according to specific patient characteristics. In particular, these analyses will focus on whether the relative therapeutic effect of ablation compared to drug therapy differs according to the following baseline variables:

- Age (<65 , 65 to 74, and ≥75 years)
- Sex (male vs. female)
- Race (white vs. racial minorities)
- AF type (paroxysmal vs. persistent, or long-standing persistent)
- Years since onset of AF (>1 vs ≤1)
- Days from most recent qualifying episode of atrial fibrillation to enrollment (>12 vs ≤12 days)
- NYHA Heart Failure Class at enrollment (no heart failure or Class I vs. ≥ Class II )
- History of congestive heart failure (yes vs. no)
- Structural heart disease (present vs. absent)
- Hypertension (present vs. absent)
- Hypertension with LVH (present vs. absent)
- CHADS2 (0 or 1 vs >1)
- $CHA_2DS_2$-VASc score (0 or 1 vs >1)
- Sleep Apnea (present vs. absent)
- Family history of atrial fibrillation (yes vs. no)
- Obesity (BMI ≥30 vs. < 30)
- Left ventricular ejection fraction (LVEF) (≤35 vs >35)
- North American vs. other international sites

These analyses will utilize the Cox model and will be accomplished by testing for interactions between the randomized treatment strategy and the specific baseline variables listed above. In addition to the formal assessment of treatment by covariate interactions, the effect of the treatment strategy characterized by a hazard ratio and 95% confidence interval will be calculated and displayed using a forest plot for the subgroups of subjects defined by the variables listed above. These descriptive hazard ratios will be carefully interpreted in conjunction with the formal interaction tests.

The effect of the treatment strategy may also be examined in other subgroups of clinical interest in addition to those listed above.

## 9. Analysis of Secondary Endpoints

### 9.1 All-cause mortality (key secondary endpoint)

***Endpoint Definition:*** Time from randomization until death.

For subjects who have not met this endpoint (and thus are censored observations), time from randomization until the subject's last follow-up contact will be used in the analysis.

Great effort will be made to gather mortality information on every randomized subject. For patients who withdraw consent or are lost to follow-up during the course of the trial, mortality databases such as the National Death Index, the Social Security Death Index, Ancestry.com, as well as any other sources of information or national registries, will be consulted in an effort to determine whether the subjects have died. Any subject found through one of these sources to have died will be considered to have met this key secondary endpoint even if not enough information is available for the CEC to classify the mode of death. Subjects for whom no vital status information can be found in these database searches will be censored at the time of their last follow-up contact.

***Statistical Methods:*** The analysis of this key secondary endpoint will use the same methods as described for the primary endpoint in Section 8. Specifically, the log-rank test will be the primary analytic tool for statistically assessing mortality differences between the two randomized treatment strategies. A hazard ratio and 95% confidence interval summarizing the difference in outcome between the two randomized arms will be computed using the Cox model. The level of significance for the assessment of this endpoint will be $\alpha$=0.05. Also, Kaplan-Meier cumulative event rates will be calculated for each randomized arm as a function of time from randomization, and the estimated event probabilities will be displayed graphically.

Supplementary analyses of this secondary endpoint will be performed using the same methods described for the primary endpoint in Section 8.1.

### 9.2 All-cause Mortality or Cardiovascular Hospitalization

***Endpoint definition:*** Time from randomization to the first event in this composite endpoint (i.e., time from randomization to either death or the subject's first cardiovascular hospitalization).

For subjects who have not experienced either component of this endpoint (i.e., who are censored observations), time from randomization until the subject's last follow-up contact will be used in the analysis.

The cause of hospitalization will be based on the *site-reported* data in the CABANA electronic data capture (EDC) system. For this endpoint, cardiovascular hospitalization will not include a hospitalization to initiate therapy. For example, among patients randomized to the ablation arm, the hospitalization to perform the ablation will not be counted for this secondary endpoint. Similarly, if a patient randomized to the drug arm is hospitalized to monitor the patient while loading the initial drug therapy, that hospitalization will not be counted. All other hospitalizations designated by the clinical site as *cardiovascular* will be counted, including the hospitalization of a patient randomized to the drug arm who is hospitalized for a therapy change (either a change to a different drug therapy or for crossing over to ablation). Hospitalization to perform second or subsequent ablation procedures will be counted as a CV hospitalization event.

***Statistical Methods:*** The analysis of this secondary endpoint will use the same methods as described for the primary endpoint and key secondary endpoint in sections 8 and 9.1, including statistical testing using the log-rank test, calculation of a hazard ratio and 95% confidence interval using the Cox model, and Kaplan-Meier cumulative event rate calculations. The level of significance for the assessment of this endpoint will be $\alpha$=0.05.

### 9.3 All-cause Mortality, Stroke, or Hospitalization for Heart Failure or an Acute Ischemic Event

***Endpoint definition:*** Time from randomization to the first event in this composite endpoint.

For subjects who have not experienced any component of this endpoint (i.e., censored observations), time from randomization until the subject's last follow-up contact will be used in the analysis. The stroke component of this endpoint is defined the same as the stroke component of the primary endpoint, namely as a disabling stroke characterized by a Rankin Scale $\geq 2$. The cause of hospitalization will be based on the *site-reported* data in the CABANA electronic data capture (EDC) system.

***Statistical Methods:*** The analysis of this secondary endpoint will use the same methods as described for the primary endpoint and secondary endpoints in the sections above, including use of the log-rank test, hazard ratio and 95% confidence interval calculated using the Cox model, and Kaplan-Meier cumulative event rate calculations. The level of significance for the assessment of this endpoint will also be $\alpha$=0.05.

This endpoint is similar to the first co-primary endpoint of the Early treatment of Atrial fibrillation for Stroke prevention Trial (EAST)[4] with the exception that EAST is using cardiovascular mortality rather than all-cause mortality as the death component of the endpoint and also a different definition of stroke (any stroke rather than a disabling stroke). Therefore, to provide results that can be compared to those of EAST, supplemental comparative analyses will be performed using cardiovascular mortality in this endpoint rather than all-cause mortality, and a definition of stroke that corresponds to the EAST definition. The statistical methods for these analyses will employ the competing risk methodology described in the next section for analyzing cardiovascular death, as well as the approach described in the preceding paragraph where all-cause mortality is a component of the endpoint.

### 9.4 Cardiovascular Death

***Endpoint definition:*** Time from randomization to the occurrence of a cardiovascular (CV) death.

Patient deaths will be classified by the Clinical Events Committee (CEC) as to whether the mode of death was due to a cardiovascular (CV) cause. If insufficient source documents are obtained to allow CEC adjudication of the cause of death, and the CEC classifies the cause of death as "unknown," then the site-reported cause of death (if available) will be used. If neither the site nor the CEC can provide a classification of the cause of death, the death will not be considered as a cardiovascular death. As supplemental analyses, however, this endpoint will also be examined using (a) only the deaths classified by the CEC as cardiovascular, and (b) using deaths

classified by the CEC as cardiovascular, but also including any deaths in the cardiovascular category that are classified as unknown by the CEC.

***Statistical Methods:*** The analysis of this endpoint will be performed using the competing risks methodology of Fine and Gray[5], where death due to a non-cardiovascular cause is considered as a competing risk. This methodology, rather than treating non-cardiovascular death as a censoring event, makes use of the cumulative incidence function, and is performed within the proportional hazards framework using the marginal failure sub-distribution associated with the event of interest (cardiovascular death). The features of the Fine-Gray approach for handling competing risks data are available as part of the PHREG procedure in SAS. The cumulative incidence function for the two arms will be estimated, statistically compared, and graphically presented using the PHREG software features in SAS. For subjects who have not died, time from randomization until the subject's last follow-up contact will be used in the analysis. The level of significance for the assessment of this endpoint will be $\alpha$=0.05.

Additional analysis of this secondary endpoint will be performed using the same methods as described for the primary endpoint and secondary endpoints in the sections above, including use of the log-rank test, hazard ratio and 95% confidence interval generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. Patients who have died but whose cause of death is classified by the CEC as non-cardiovascular will be censored at the time of their non-CV death. The number of such patients is expected to be relatively small in this study. The results of these two different approaches will be carefully examined and synthesized into an overall interpretation of treatment differences with respect to CV death.

### 9.5 Cardiovascular Death or Disabling Stroke

***Endpoint definition:*** Time from randomization to the occurrence of either a cardiovascular death or disabling stroke.

All potential strokes will be carefully reviewed and classified as disabling by the neurological events committee using the Rankin Stroke Scale as described in Section 8. The neurological committee that reviews and adjudicates the strokes is an important part of the overall clinical events adjudication process. A disabling stroke will be defined as a stroke whose severity on the Rankin Scale is $\geq 2$. Cardiovascular deaths will be classified as described in Section 9.4 above.

***Statistical Methods:*** The analysis of this secondary endpoint will also use the competing risks methodology of Fine and Gray[5] described in Section 9.4 for the secondary endpoint of CV death where death due to a non-cardiovascular cause is considered as a competing risk rather than a censoring event. The analysis will make use of the cumulative incidence function, and (as described in Section 9.4) will be performed within the proportional hazards framework using the marginal failure sub-distribution associated with the event of interest. The cumulative incidence function for the two arms will be estimated, statistically compared, and graphically presented using the PHREG procedure in SAS. For subjects who have not died, time from randomization until the subject's last follow-up contact will be used in the analysis. The level of significance for the assessment of this endpoint will be $\alpha$=0.05.

As in the case of the cardiovascular death endpoint, additional analysis of this secondary endpoint will be performed using the log-rank test, hazard ratio and 95% confidence interval generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. Patients who have died but whose cause of death is classified by the CEC as non-cardiovascular will be censored at the time of their non-CV death. Results of these two different approaches will be carefully compared and synthesized into an overall interpretation of treatment differences with respect to the endpoint of CV death or disabling stroke.

### 9.6 Arrhythmic Death or Cardiac Arrest

***Endpoint definition:*** Time from randomization to the occurrence of either an arrhythmic death or cardiac arrest.

Since cardiac arrest is a component of the primary endpoint, all potential cardiac arrests will be reviewed and adjudicated by the Clinical Endpoints Committee. As described earlier, the mode of death is also classified by the Clinical Endpoints Committee.

***Statistical Methods:*** The analysis of this secondary endpoint will proceed in much the same way as described above for the endpoint of CV death or disabling stroke and use the competing risks methodology of Fine and Gray[5] described in previous sections. In this analysis, death due to a cause other than arrhythmic is considered as a competing risk rather than a censoring event. The analysis will make use of the cumulative incidence function, and will be performed within the proportional hazards framework using the marginal failure sub-distribution associated with the event of interest. The cumulative incidence function for the two arms will be estimated, statistically compared, and graphically presented using the features available in the PHREG procedure in SAS. For subjects who have not died, time from randomization until the subject's last follow-up contact will be used in the analysis. The level of significance for the assessment of this endpoint will be $\alpha$=0.05.

For this endpoint, we will also perform additional analysis using the log-rank test, hazard ratio and 95% confidence interval generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. Patients who have died but whose cause of death is classified by the CEC as other than arrhythmic will be censored at the time of the non-arrhythmic death. The results of these two different approaches will be compared and synthesized into an overall interpretation of treatment differences with respect to the endpoint of arrhythmic death or cardiac arrest.

### 9.7 Heart Failure Death

***Endpoint definition:*** Time from randomization to the occurrence of a heart failure death.

Classification of deaths as being due to heart failure will be performed by the blinded Clinical Endpoints Committee.

***Statistical Methods:*** The analysis of this endpoint will make use of the competing risks methodology of Fine and Gray[5] described in previous sections for other secondary endpoints. In this case, death due to a cause other than heart failure will considered as a competing risk. The

features of the Fine-Gray approach for handling competing risks data available as part of the PHREG procedure in SAS will be employed to estimate the cumulative incidence function for each of the two arms, to statistically compare these functions, and graphically present them using the software features in SAS. The level of significance for the assessment of this endpoint will also be $\alpha$=0.05.

Additional analysis of this secondary endpoint will also be performed using the log-rank test, hazard ratio and 95% confidence interval (CI) generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. For this analysis, patients who have died, but their death is classified by the CEC as due to a cause other than heart failure, will be censored at the time of their non-heart failure death. The results of the competing risk analysis and of this traditional analysis will be compared to synthesize an overall interpretation of the treatment differences.

### 9.8 Recurrent Atrial Fibrillation

***Endpoint definition:*** This particular outcome will be analyzed with different endpoint definitions in order to provide an overall perspective of the extent of the AF burden experienced by the study subjects. The first analysis will consider the first episode of recurrent AF as the endpoint, defined as an episode $\geq$30 seconds following the end of the 90-day blanking period, where the 90-day blanking period begins at the time of receiving the randomly assigned treatment (either ablation or drug therapy). The endpoint will be defined as the time until the first recurrence of AF according to this definition. Similar analyses will be performed where the endpoint is defined as: (a) an episode of *symptomatic* recurrent AF, (b) an episode of atrial fibrillation or atrial flutter or atrial tachycardia, and (c) an episode of *symptomatic* atrial fibrillation, atrial flutter, or atrial tachycardia.

***Analysis Population:*** Because recurrent atrial fibrillation is defined with respect to the end of the blanking period, which begins after receipt of the assigned treatment, all analyses for recurrent atrial fibrillation will include only patients who have received their randomly assigned treatment and live to the end of the blanking period.

In addition, separate analyses will be performed according to the method used for monitoring AF recurrence. As indicated in Section 6 (Data Sources), rhythm data for many of the patients enrolled in CABANA (approximately 2/3 of the randomized patients) are collected using the "CABANA Box," a robust rhythm monitoring system provided by Medicomp. The rhythm monitoring data are sent from Medicomp to the Duke Clinical Research Institute (DCRI), where a set of programmed rules are run to determine which rhythms should be submitted to the ECG core lab for review and adjudication. Following core lab review, the adjudicated results are electronically transmitted back to the DCRI. Sites that are not using the "CABANA Box" (generally these are international sites where use of the CABANA Box was not approved) submit rhythm data through a secure portal using the CABANA website, which is then processed and submitted to the core lab for adjudication. Because the CABANA Box data are more comprehensive and more routinely and rigorously collected, separate analyses of recurrent atrial arrhythmias will be performed using (a) patients whose rhythms were monitored using the CABANA Box, (b) all remaining patients (i.e., non-CABANA Box patients), and (c) the pooled group of all patients (with use of CABANA Box as a covariate if justified). In the case of

patients who initially used the CABANA Box and later ceased to provide information via CABANA Box, the analysis will censor that data at the time of last CABANA Box recording. The primary analysis of this endpoint will be the analysis of patients with CABANA Box data because of the more comprehensive collection of that rhythm monitoring information. Analysis which includes non-CABANA Box patients ("b" and "c" above) will be considered supplementary analysis.

***Statistical Analysis:*** First, descriptive summaries of baseline characteristics of subjects receiving their assigned treatment and living to the end of the blanking period in both treatment arms will be presented to determine how comparable the patients in the two arms are for this analysis. Continuous variables will be presented in terms of percentiles (median, $25^{th}$ and $75^{th}$ percentiles), while discrete variables will be summarized in terms of frequencies and percentages.

The treatment arms will be compared using statistical methods described in the sections above for other secondary endpoints, including the competing risk method of Fine and Gray[4], as well as use of the log-rank test, hazard ratio and 95% confidence interval generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. In this case, all-cause mortality will be a competing risk and zero-time (i.e., when the clock starts) for this analysis will be the end of the 90-day blanking period. Analyses will be repeated with adjustment for the baseline covariates described in section 8.1.

Of key interest will be whether the effect of ablation on recurrent atrial arrhythmias is similar or different according to the type of AF the patients experienced prior to enrolling in the trial (i.e., whether the effect varies according to whether patients' baseline AF was paroxysmal, persistent or long-standing persistent). This analysis will utilize the Cox model to test for an interaction between the randomized treatment strategy and the baseline type of AF classified as (a) paroxysmal or (b) persistent or longstanding persistent. In addition to the interaction test, the effect of the treatment strategy characterized by a hazard ratio and 95% confidence interval and Kaplan-Meier event rate calculations will be calculated and displayed for the subgroups of patients whose baseline AF history was (a) paroxysmal and (b) persistent or longstanding persistent. Similar comparative analyses will be performed in subgroups defined by age ($<65$, 65-74, and $\geq 75$) and by $CHA_2DS_2$-VASc Score (0 or 1 vs >1).

### 9.9 Cardiovascular Hospitalization

A key part of the analysis and interpretation of the results of this endpoint will be the analysis performed in connection with secondary endpoint #2 (the composite of death or cardiovascular hospitalization). Analysis of cardiovascular hospitalization as an endpoint by itself must be analyzed and interpreted cautiously, particularly if there is a mortality difference between the two arms of the trial. If patients in one arm have lower mortality, that arm may have more hospitalizations simply because more patients are alive and at risk for needing to be hospitalized. Another important aspect of the analysis of this endpoint is that patients may have multiple hospitalizations, and so a comprehensive analysis of hospitalizations must take into account the recurrent hospitalization events rather than simply the first one. Consequently, there will be

several parts to the analysis of this endpoint in addition to a careful examination of the results of the composite endpoint consisting of death or cardiovascular hospitalization (see Section 9.2).

***Statistical Methods:***

a. The first analysis will consider time until the first cardiovascular hospitalization in each arm and use the competing risks methodology of Fine and Gray[5] where death from any cause is considered as a competing risk. The analysis will make use of the cumulative incidence function, and will be performed within the proportional hazards framework using the approach described in Section 9.4. The cumulative incidence function for each of the two arms will be estimated, statistically compared, and graphically presented using the PHREG procedure in SAS. For subjects who have not experienced a CV hospitalization, time from randomization until the subject's last follow-up contact will be used in the analysis. The level of significance for the assessment of this endpoint will be $\alpha$=0.05.

b. An additional analysis of this endpoint will be performed using the log-rank test, hazard ratio and 95% confidence interval generated using the Cox model, and Kaplan-Meier cumulative event rate calculations. Patients who have died will be censored at the time of their death. This analysis and the approach in (a) above will be compared and summarized into an overall interpretation of treatment differences with respect to the endpoint of time to the first CV hospitalization.

c. The analyses described above will be extended to a multiple events analysis, designed to take into account multiple hospitalization events that may occur during the course of a patient's follow-up. When considering one event per patient as in (a) or (b) above (i.e., the time until the first event), those events are independent because they occur in different patients. When considering multiple events for a patient, the events are no longer all independent because some of them are occurring within the same patient. This correlation (covariance) must be taken into account in deriving inferences. The multiple-event methods that will be used represent different extensions of Cox model methodology, and take into account the timing of the events and the correlation among multiple events occurring within an individual patient.

   The main statistical approach will use a generalization of the Cox model developed by Andersen & Gill[6,7], with robust standard errors to account for heterogeneity and for the correlation between recurrent events within a patient.

   Similar to time-to-1st event methods, a terminal event (death) could introduce bias in the recurrent events analysis, especially if there is a differential mortality between the study arms. Thus, the Ghosh and Lin[8] nonparametric method can be utilized in calculating the cumulative rate of CV hospitalizations, while accounting for mortality as a competing risk. In addition, the average number of CV hospitalizations per 100-person years utilizing the Ghosh-Lin method can be calculated for each treatment group.

   The Andersen-Gill recurrent events method will be performed using PROC PHREG in SAS software. However, the Ghosh-Lin method requires implementation of a validated package in R software.

### 9.10 Medical Costs, Resource Use and Cost Effectiveness.

Cost data will be analyzed by the CABANA EQOL Coordinating Center in close collaboration with the Data Coordinating Center, and the analysis approach will be detailed in a separate analytic plan.

### 9.11 Quality of Life.

Quality of life data will also be analyzed by the CABANA EQOL Coordinating Center in close collaboration with the Data Coordinating Center. Again, the analysis approach will be detailed in a separate analytic plan.

### 9.12 Major Non-Endpoint Adverse Events.

Major adverse events (AE) will be monitored and reported during the trial. For drug therapy, the major serious or life-threatening events include unanticipated hospitalizations for heart failure, drug-related pro-arrhythmic events, and pulmonary toxicity from amiodarone. For the ablation arm, the major events include symptomatic pulmonary vein stenosis requiring intervention, myocardial perforation/tamponade requiring intervention, and atrial-esophageal fistula formation. The frequency of each of these events (and other serious adverse events such as peripheral thromboembolic events) will be descriptively tabulated using MedDRA System Organ Class. Treatment comparisons with respect to these serious adverse events will be more descriptive rather than statistically rigorous (since different adverse events are expected in the two arms), bringing to bear clinical judgment as to the relative seriousness of these various adverse events. Other adverse complications or adverse events of either therapy will be tabulated and descriptively summarized, using statistical comparisons of the treatment groups based on chi-square tests where appropriate, and interpreting such comparisons in the context of treatment differences in the primary and major secondary clinical endpoints.

### 9.13 LA Size, Morphology and Function.

For the subjects participating in the CT/MR imaging sub-study, left atrial size, morphology, and function will be examined. Of interest will be the change in atrial size (volume) and function (i.e., atrial ejection fraction) from the pre-treatment to the post-treatment study within each treatment arm, and then importantly, a comparison of whether the change in size and function in the patients randomized to ablation therapy differs from the change in size and function of the patients randomized to drug therapy.

The distribution of these measures from the pre- and post-treatment studies in each treatment arm will be descriptively characterized using summary descriptive statistics and graphical tools including box and whisker plots. The statistical significance of the changes in left atrial volume and atrial ejection fraction from pre- to post-treatment studies within each treatment arm will be assessed using general linear models (GLM), including the baseline value as a covariate, and assessing treatment group differences. If there appears to be differences in other baseline

covariates between groups, adjustment for covariates will be considered. Secondary analyses will use a paired-sample t-test instead of GLM.

In dealing with this type of data, the completeness of which is dependent on obtaining both the pre- and post-treatment studies, there will inherently be missing data resulting from patients in whom it was not possible to obtain both studies (because they died or failed to return for other reasons to enable the follow-up measurements to be made). The completeness of the data will be fully described, and the statistical comparisons will be based on patients for whom both the baseline and follow-up studies are available.

## 10. "On-Treatment" Analysis

The intention-to-treat analyses in this trial will constitute the primary analyses and will serve as the standard for interpreting treatment differences in the key clinical outcomes. However, because a number of patients in the drug arm may cross over to receive ablation during the trial, and some patients randomized to ablation may not undergo the procedure, we will supplement the intent-to-treat comparisons of the primary endpoint and selected secondary endpoints with "on-treatment" comparisons. The "on-treatment" analysis will involve a comparison of patients who received ablation (even if originally assigned to the drug arm) versus those who did not. Statistically, this will be accomplished using the Cox proportional hazards model with ablation included in the analysis as a time-dependent covariate. This analysis will be repeated for the primary endpoint, all-cause mortality, all-cause mortality or cardiovascular hospitalization, all-cause mortality, stroke, or hospitalization for heart failure or acute ischemic events, cardiovascular death, cardiovascular death or disabling stroke, arrhythmic death or cardiac arrest, heart failure death, and cardiovascular hospitalization with and without adjustment for covariates listed in section 8.1. No analyses will include competing risks, and on-treatment analysis will not be conducted for multiple events analyses (section 9.9). For recurrent atrial fibrillation, we will only conduct this on-treatment analysis for the first episode of recurrent AF without competing risks in the 3 patient populations (Cabana box, non-cabana box, and pooled). Secondary analyses may consider other measures of recurrent atrial fibrillation (e.g., *symptomatic* recurrent AF, atrial fibrillation or atrial flutter or atrial tachycardia, and *symptomatic* atrial fibrillation, atrial flutter, or atrial tachycardia).

The time-dependent ablation variable to be included in the analysis will be created using 3 different methods: For <u>method 1</u>, the ablation variable starts as 0 (no ablation) for all patients. For patients who received ablation, the ablation variable is set to 1 at the time of ablation. In <u>method 2</u>, the ablation variable starts as 1 for all patients who were randomized to ablation and actually received ablation. For all other patients, the ablation variable is started as 0 and set to 1 at the time of ablation if they received ablation during the trial. <u>Method 3</u> has the ablation variable started as 1 for patients who were randomized to ablation. For all other patients (i.e., randomized to drug), the ablation variable is started as 0 and set to 1 at time of ablation.

Among the 3 methods described above, method 2 will be considered the main statistical method in handling cross-overs via a time-dependent variable in Cox model. The other 2 methods are considered as sensitivity analyses.

Differential treatment effects for the primary endpoint and major secondary endpoints of interest using on-treatment analyses will also be conducted. Subgroups of interest will include those listed in section 8.1.

Because treatment assignment is no longer random, results can be biased by an association between the likelihood of treatment change and the risk of a clinical event (i.e., factors that make a patient more likely to cross over may also make them more likely to have an event). Thus, such analyses will be cautiously interpreted. Each of the analyses described in this section will be covariate adjusted using the covariates described in section 8.1 for the primary endpoint.

## 11. Per-Protocol Analysis

Using the methods described for the primary endpoint in Section 8, an analysis will also be performed to compare the primary endpoint and the key secondary clinical outcomes of the two treatment strategies among the subset of patients who fully satisfied the inclusion/exclusion criteria and received the treatment to which they were randomly allocated. This analysis will include patients randomized to the drug arm who were treated with drug therapy, and patients randomized to the ablation arm who underwent the ablation within 6 months following enrollment in the trial. The follow-up of drug-arm patients who crossed over to ablation will be censored at the time of the ablation. Patients randomized to the ablation arm who were not ablated within 6 months will be excluded from this analysis. Additional sensitivity analyses will be performed in which the window for ablation is varied to include a shorter interval (3 months) and a longer window (1 year).

The analyses outlined in Sections 10 and 11 will be strictly supplementary to the intent-to-treat analyses described earlier.

## 12. Multiple Comparisons

With the primary hypothesis and the various secondary endpoints that have been outlined, there is a multiplicity of analyses to be performed, which leads to an increased probability that at least one of the comparisons could be "significant" by chance. There are adjustments (e.g., based on the Bonferroni inequality) that could be used to preserve the overall type I error level. To attempt to adjust for the effects of the repeated significance testing that will occur as part of the interim monitoring (discussed below), plus adjust for the multiplicity of secondary endpoints, would require that very small significance levels be used for every comparison. The overall level of significance will be 0.049 for the assessment of the primary endpoint (see Section 13) and 0.05 for all secondary endpoints. However, to account in part for the multiplicity of comparisons involving secondary endpoints, we will be conservative in the interpretation of the analyses, taking into account the degree of significance, and looking for consistency across endpoints. The actual p-value for each comparison will be reported to aid in the overall interpretation. We have

also pre-specified the primary and secondary outcome variables to avoid over-interpretation of strictly exploratory comparisons.

## 13. Interim Analyses

For ethical reasons, an interim examination of key safety and endpoint data has been performed at regular intervals during the course of the trial. The primary objective of these analyses was to evaluate the accumulating data for an unacceptably high frequency of negative clinical outcomes in either of the treatment arms. In addition, the interim monitoring involved a review of the aggregated event rates, patient recruitment, compliance with the study protocol, status of data collection, and other factors which reflected the overall progress and integrity of the study. The results of the interim analyses and status reports were carefully and confidentially reviewed by an NHLBI-appointed Data and Safety Monitoring Board (DSMB).

The DSMB met at approximately 6-month intervals to review the accumulating data. Prior to each meeting, the Statistical and Data Coordinating Center (SDCC) at the DCRI conducted the desired statistical analyses and prepared a summary report for review by the DSMB. The extracted data files and analysis programs for each DSMB report were archived and maintained at the DCRI. Reports were presented describing the progress of patient enrollment, the rates of compliance with therapy, and the frequency of protocol violations. The SDCC also reported on the number (status) of data forms completed and the overall quality and completeness of the information.

In CABANA, a group sequential method similar to that proposed by O'Brien and Fleming[9] was used as a guide in interpreting interim analyses. This procedure required large critical values early in the study, but relaxed (i.e., decreased) the critical value as the trial progressed. Because of the conservatism early in the trial, the critical value at the final analysis is near the "nominal" critical value. The actual method used for the interim monitoring was the Lan and DeMets[10] group sequential method with monitoring boundaries based on a two-sided, symmetric O'Brien-Fleming type spending function. The Lan-DeMets approach only required specification of the rate at which the Type I error (which in this trial is $\alpha = 0.05$ for the primary endpoint) was "spent". The procedure allowed "spending" a portion of $\alpha$ at each interim analysis in such a way that at the end of the study, the total Type I error did not exceed 0.05. The specific formulation of the spending function is described in Chapter 5 of Proschan et al (formula 5.4)[11]. Since the number of looks and the increments between looks did not need to be predetermined, it allowed considerable flexibility in the monitoring process for accommodating comparative examinations of the data in response to concerns of the DSMB that may arise during the course of the trial.

The analytic approach used at the interim analyses for assessing treatment differences consisted of the time-to-event analysis methods described previously, except that interpretation of statistical significance associated with treatment comparisons of key study endpoints was guided using the group sequential monitoring boundaries. The critical value of the test statistic and the

corresponding p-value required for significance was presented so that significance could be assessed precisely.

Although the CABANA DSMB has met approximately twice each year, the board has seen interim treatment-specific primary endpoint data only one time during the course of the trial. At all other meetings, only overall (aggregated) event data have been presented to them. As a result of the single treatment-specific DSMB presentation, the overall level of significance for the final analysis of the primary endpoint has been calculated to be .049 (using the O'Brien-Fleming boundary).

## References


1. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* 2nd edition. 2002, John Wiley & Sons, Inc. Hoboken, New Jersey.
2. Cox DR. Regression models and life-tables (with discussion). *J Royal Statist Soc B* 1972; 34:187-220.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Statist Assn* 1958; 53:457-481.
4. Kirchhof P, Breithardt G, Camm AJ, Crijns HJ, Kuck K-H, Vardas P, Wegscheider K. Improving outcomes in patients with atrial fibrillation: Rationale and design of the Early treatment of Atrial fibrillation for Stroke prevention Trial. *Am Heart J* 2013; 166:442-448.
5. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Statist Assoc* 1999; 94:496-509.
6. Anderson PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat* 1982; 10:1100-1120.
7. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc* 1989; 84:1074-1078.
8. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics* 2000; 56:554-62.
9. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; 35:549-56.
10. Lan KK, DeMets L. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70:659-63.
11. Proschan MA, Lan KKG, Wittes JT. *Statistical Monitoring of Clinical Trials – A Unified Approach*. 2006