Statistical Analysis Plan

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation

OPTION

APPROVED ELECTRONICALLY ON 12-JAN-2024

BSC Project Number: S2239

NCT #03795298

TABLE OF CONTENTS

1 PROTOCOL SYNOPSIS

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (**OPTION**)	
Study Objective(s)	The primary objective of this study is to determine if left atrial appendage closure with the WATCHMAN FLX Device is a reasonable alternative to oral anticoagulation following percutaneous catheter ablation for high risk patients with non-valvular atrial fibrillation.
Planned Indication(s) for Use	The planned indication for use within the OPTION study is as follows: The WATCHMAN and WATCHMAN FLX Device are indicated to reduce the risk of thromboembolism from the left atrial appendage in subjects with non-valvular atrial fibrillation who: • Are at increased risk for stroke and systemic embolism based on CHA_2DS_2-VASc scores following catheter ablation of atrial fibrillation; and • Are deemed to be suitable for anticoagulation therapy Note: In countries where CE mark applies, the OPTION indication is within the CE mark approved Indication for Use.
Study Design	This study is a prospective, randomized, multi-center, global investigation to determine if left atrial appendage closure with the WATCHMAN FLX Device is a reasonable alternative to oral anticoagulation in patients after AF ablation.
Planned Number of Subjects	A maximum of 1600 subjects will be randomized in the study. *Note the total number of enrolled patients is expected to exceed the number of randomized subjects since sites without WATCHMAN FLX experience are required to perform two roll-in cases. A maximum of 260 patients will be treated in the roll-in phase of the study, including approximately 130 roll-in subjects in the United States.* *The roll-in cohort will be analyzed separately from the primary cohort of randomized subjects.*
Planned Number of Centers / Countries	Up to 150 investigational centers worldwide
Primary Effectiveness Endpoint	WATCHMAN therapy is non-inferior for the occurrence of stroke (including ischemic and/or hemorrhagic), all cause death, and systemic embolism at 36 months.

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (**OPTION**)	
Primary Safety Endpoint	WATCHMAN therapy is superior for non-procedural bleeding through 36 months (ISTH major bleeding and clinically relevant non-major bleeding)
Secondary Endpoint	WATCHMAN therapy is non-inferior for ISTH major bleeding at 36 months (including procedural bleeding)
Additional Analysis	The occurrence of: • Stroke • Ischemic stroke • Hemorrhagic stroke • Disabling stroke • Non-disabling stroke • Systemic embolism • Procedural and non-procedural bleeding • All-cause death • Cardiovascular/unknown death • Non-cardiovascular death • Device related Thrombus • Device Seal • Single procedure freedom from AF • Healthcare resource utilization • Quality of life
Method of Assigning Patients to Treatment	A subject who signs informed consent is considered enrolled in the study. Subjects will be randomized to OAC or WATCHMAN FLX in equal fashion. Randomization will be stratified by sequential vs. concomitant planned ablation to help ensure balance of treatment assignments within the sequential and concomitant groups.
Follow-Up Schedule	Study procedures and follow-up visits will occur as follows: • Consent – Must be obtained within 30 days prior to randomization • Randomization ○ Prior ablation (Sequential group) – randomization must be performed between 90 and 180 days after the most recent AF ablation procedure. WATCHMAN FLX Implant – must be performed within 10 days following randomization if subject is randomized to the device arm

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (**OPTION**)	
	○ Planned prospective ablation (Concomitant group) – the ablation +/− WATCHMAN FLX Implant must be performed within 10 days of randomization • 3 Month Follow-up (90 ± 15 days from randomization) • 12-Month Follow-up (365 ± 30 days from randomization) • 24-Month Follow-up (730 ± 60 days from randomization) • 36-Month Follow-up (1095 + 60 days from randomization) – this follow-up must occur on or after 1095 days and on or before 1155 days Note: for Roll-in subjects the date of implant will be used to calculate Follow-up windows instead of the date of randomization.
Study Duration	The duration of the study is expected to last approximately 64 months. The duration of individual subject participation is expected to last approximately 36 months but may vary per subject.
Control Group Medication Therapy	Following randomization, control subjects must continue or start market-approved OAC used per IFU for atrial fibrillation stroke prevention and should remain on it for the duration of the trial.
Device Group Medication Therapy	After the WATCHMAN FLX implant, Device Group subjects will be prescribed market-approved OAC and aspirin (75-100mg recommended) until the 3-month visit followed by aspirin until at least the 12-month visit (recommended for duration of the trial).
Test Device and sizes	The WATCHMAN FLX Left Atrial Appendage Closure Device with Delivery System (consisting of the Delivery Catheter with a pre-loaded Closure Device) WATCHMAN FLX is available in 20, 24, 27, 31, and 35mm models to fit left atrial appendage ostia widths ranging from 14.0 - 31.5mm.
Inclusion Criteria	1. The subject is of legal age to participate in the study per the laws of their respective geography. 2. Underwent a prior catheter ablation procedure for non-valvular AF between 90 and 180 days prior to randomization (sequential) or is planning to have clinically indicated catheter ablation within 10 days of randomization (concomitant). 3. The subject has a calculated CHA2DS2-VASc score of 2 or greater for males or 3 or greater for females. 4. The subject is deemed by the treating physician to be suitable for the protocol defined pharmacologic regimens. 5. The subject is able to undergo TEE examinations.

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (**OPTION**)	
	6. The subject or legal representative is able to understand and is willing to provide written informed consent to participate in the trial. 7. The subject is able and willing to return for required follow-up visits and examinations.
Exclusion Criteria	1. The subject is currently enrolled in another investigational study that would directly interfere with the current study, except when the subject is participating in a mandatory governmental registry, or a purely observational registry with no associated treatments. Each instance must be brought to the attention of the sponsor to determine eligibility, regardless of type of co-enrollment being proposed. 2. The subject requires long-term anticoagulation therapy for reasons other than AF-related stroke risk reduction, for example due to an underlying hypercoagulable state (i.e., even if the device is implanted, the subjects would not be eligible to discontinue OAC due to other medical conditions requiring chronic OAC therapy). 3. The subject is deemed by the treating physician to be unsuitable for chronic anticoagulation and/or aspirin therapy due to bleeding risk, allergy, or other reasons. 4. The subject had or is planning to have any cardiac or major non-cardiac interventional or surgical procedure (excluding non-valvular AF ablation and cardioversion) within 30 days prior to or 60 days after randomization [including, but not limited to: percutaneous coronary intervention (PCI), other cardiac ablation (VT ablation, etc.), etc.]. 5. The subject had a stroke or transient ischemic attack (TIA) within the 60 days prior to randomization. 6. The subject had a prior major bleeding event per ISTH definition within the 14 days prior to randomization. Lack of resolution of related clinical sequelae, or planned and pending interventions to resolve bleeding/bleeding source, are a further exclusion regardless of timing of the bleeding event. 7. The subject has had a myocardial infarction (MI) documented in the clinical record as either a non-ST elevation MI (NSTEMI) or as an ST-elevation MI (STEMI), with or without intervention, within 90 days prior to randomization. 8. The subject has a history of atrial septal repair or has an ASD/PFO device. 9. The subject has an implanted mechanical valve prosthesis in any position.

Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (**OPTION**)	
	10. The subject is of childbearing potential and is, or plans to become pregnant during the time of the study (method of assessment upon study physician's discretion) 11. The subject has a documented life expectancy of less than two years. 12. The subject has a cardiac tumor. 13. The subject has signs/symptoms of acute or chronic pericarditis. 14. There is evidence of tamponade physiology. 15. The subject has contraindications (anatomical or medical) to percutaneous catheterization procedures. 16. The subject has documented NYHA Class IV heart failure. 17. The subject has documented surgical closure of the left atrial appendage. 18. The subject has an active infection.
Transthoracic Echo Exclusion Criteria	1. The subject has LVEF < 30% 2. The subject has an existing pericardial effusion with a circumferential echo-free space > 5mm. 3. The subject has a high- risk patent foramen ovale (PFO) with an atrial septal aneurysm excursion > 15mm or length > 15mm. 4. The subject has a high-risk PFO with a large shunt defined as early, within 3 beats and/or substantial passage of bubbles. 5. The subject has significant mitral valve stenosis (i.e., MV area <1.5 cm^2). Note: Criteria obtained from cardiac imaging performed within 180 days prior to randomization may be used if all the exclusion criteria can be evaluated.
Multiple Interventions During Index Procedure	Percutaneous catheter ablation using currently available non-surgical standard techniques and market-approved technology may be performed at the time of the WATCHMAN FLX implant procedure **only for Concomitant Ablation subjects** • The procedures must occur on the same day with the ablation occurring prior to the WATCHMAN FLX implant. • Catheter ablation of the LAA, non-standard ablation techniques (e.g. CFAEs and hybrid ablation) and any non-AF-related ablation (e.g. VT) are not permitted. • Other concomitant procedures are also not permitted, including, but not limited to, transcutaneous valve procedures, pacemaker or ICD generator change, etc.

<table>
<tr><th colspan="2">Comparison of Anticoagulation with Left Atrial Appendage Closure after AF Ablation (OPTION)</th></tr>
<tr><td></td><td>Note: Sequential Ablation subjects who need a repeat ablation cannot combine it with the WATCHMAN FLX implant procedure.
For all patients, cavotricuspid isthmus (CTI) ablation and cardioversion may be performed during ablation.</td></tr>
<tr><td>Statistical Test Method</td><td>Hypotheses testing in this study will use standard statistical methodology. Each primary endpoint will be assessed vs. a performance goal. To declare success, the primary endpoints must be met. The details of each endpoint analysis are listed in the table below.
<table>
<tr><th>Endpoint</th><th>Assessment period</th><th>Expected Rate</th><th>Non-Inferiority Margin</th><th>Test (α=2.5%) NI (α=5%) Sup</th><th>Expected Attrition</th><th>Power subjects</th></tr>
<tr><td>Primary Efficacy Endpoint</td><td>36-month follow-up visit</td><td>10% in both arms</td><td>5%</td><td>One-sided NI Test of KM event rates</td><td>15%</td><td>84% (N=1600)</td></tr>
<tr><td>Primary Safety Endpoint</td><td>36-month follow-up visit</td><td>Device: 14%
Control: 20%</td><td>NA</td><td>Two-sided log-rank test</td><td>20%</td><td>>86.4% (N=1600)</td></tr>
<tr><td>Secondary Endpoint</td><td>36-month follow-up visit</td><td>Device: 11%
Control: 11%</td><td>5.25%</td><td>One-sided NI Test of KM event rates</td><td>20%</td><td>84.3% (N=1600)</td></tr>
</table>
NI-Non-Inferiority, Sup-Superiority</td></tr>
</table>

2 INTRODUCTION

This statistical analysis plan (SAP) addresses the planned analyses for the OPTION trial based on the approved protocol version E. The primary efficacy endpoint analysis will use time-to-event Kaplan-Meier estimates and Greenwood formula for variance to assess noninferiority of the Device vs the Control at a specific timepoint. The primary safety endpoint analysis will use log-rank statistics to assess superiority of Device vs Control over time.

3 ENDPOINT ANALYSES

3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the occurrence of stroke (including ischemic and/or hemorrhagic), all cause death, and systemic embolism at 36 months. This endpoint is defined as the Kaplan Meier estimate of time to first occurrence of stroke (including ischemic and/or hemorrhagic), all cause death, or systemic embolism at 36 months and will be tested for non-inferiority of WATCHMAN to control.

3.1.1 Hypotheses

The objective of the primary efficacy endpoint analysis is to test the null hypothesis that the difference in cumulative incidence between the Device and Control groups is greater than a prespecified noninferiority margin δ. The null hypothesis will be tested vs. the one-sided alternative hypothesis that the difference in cumulative incidence is less than a noninferiority margin δ.

H_0: $S_1(t) \geq S_0(t) + \delta$

H_a: $S_1(t) < S_0(t) + \delta$,

where $S_1(t)$ and $S_0(t)$ are the Kaplan Meier estimates for the cumulative incidence of stroke (including ischemic and/or hemorrhagic), all-cause death, and systemic embolism at 36 months for the Device and Control groups, respectively.

3.1.2 Sample Size for the Primary Efficacy Endpoint

The power and sample size were calculated employing Farrington Manning methods for non-inferiority using SAS version 9.4. The assumptions and parameters pertaining to this design are as follows:

- Power = 85.9%
- Alpha = 0.025 (one-sided)
- Expected Device and Control group performance: cumulative incidence rate of 10% at 36 months in both groups
- A noninferiority margin δ = 5%
- Attrition: cumulative attrition rate of 15% in the Device Group and in the Control group at 36 months.

The expected cumulative incidence of 10% at 36 months in the Device group was derived from historical event rates from the PROTECT AF, CAP, PREVAIL, and CAP2 studies. WATCHMAN subjects with previous ablations from these studies experienced a 14.4% event at

3 years. The cumulative incidence is expected to be lower in the OPTION study due to enhanced implant experience over time, device improvements, and a decreased baseline risk profile for subjects enrolled in the OPTION study. The expected cumulative incidence in the Control arm is expected to be equal to that of the Device arm.

The non-inferiority margin of 5% represents a relative risk of 1.5.

Given the above assumptions, 1360 subjects will be required. In order to account for up to 15% expected rate of attrition which includes subjects randomized but not treated and subjects withdrawn or lost-to-follow-up through 36 months, a total of 1600 (1360/(1-0.15)) subjects need to be enrolled.

A sample size of 1600 subjects provides 85.9% power for the Primary Effectiveness Endpoint analysis. Therefore, a maximum sample size of 1600 is determined.

In addition, the power and sample sizes have been re-evaluated for the non-inferiority testing settings that would match the method described to perform the hypothesis analysis, namely, differences in the Kaplan-Meier event rates of the respective endpoints using the Greenwood formula.

The operating characteristics of the statistical test for the primary effective endpoint, and the secondary safety endpoint, are calculated by simulating 10,000 trials, for each endpoint, for a given sample size using custom-written code in SAS. The results are consistent with the sample size calculation performed using a parametric formula, namely the Farrington-Manning method above. The table below presents the simulation results for primary efficacy endpoint.

Power and type I error of the Simulation for Primary Effective Endpoint

Method	**Farrington-Manning**	**Simulation**
Power	85.9%	89.3%
Alpha (Type I error)	0.025	0.024

3.1.3 Statistical Methods for the Primary Efficacy Endpoint

The primary efficacy endpoint event rate from the WATCHMAN FLX Device and Control arms, $S_1(t)$ and $S_0(t)$, respectively, will be estimated by the Kaplan Meier method. The timepoint of 36 months is defined as the 1095^{th} day post-randomization. The 97.5% one-sided upper bound of confidence limit of the difference between WATCHMAN FLX and Control rates will be calculated using Greenwood formula for the variance of the Kaplan Meier estimates. The objective is met if this confidence limit is less than the predefined noninferiority margin of 5%.

3.1.4 Superiority Test for the Primary Efficacy Endpoint

If the non-inferiority hypothesis objective is met and the observed event rate favors the Device arm, i.e., $S_1(t) < S_0(t)$, a superiority test will be subsequently performed and a one-sided p-value of 0.025 will be considered significant. No multiplicity adjustment is needed[1].

[1] B Logan and A Tamhane, Superiority Inferences on Individual Endpoints Following Noninferiority Testing in Clinical Trials, Biometrical Journal 50 (2008) 5.

3.2 Primary Safety Endpoint

The primary safety endpoint is non-procedural bleeding through 36 months (ISTH major bleeding and clinically relevant non-major bleeding). This endpoint is defined as the Kaplan Meier estimate of time to first occurrence of non-procedural ISTH major bleeding or clinically relevant non-major bleeding through 36-months. Non-procedural events are those occurring after 3 days, calculated from implant or attempted implant datefor Device patients and from date of randomization for Control patients. This endpoint will be tested for superiority of WATCHMAN vs Control.

3.2.1 Hypotheses

The objective of the primary safety endpoint analysis is to test the null hypothesis that time-to-event distributions do not differ between the Device and Control groups. The null hypothesis will be tested vs. the 2-sided alternative hypothesis that the time-to-event curves are different.

H_0: $S_1(t) = S_0(t)$

H_a: $S_1(t) \neq S_0(t)$,

where $S_1(t)$ and $S_0(t)$ are the time-to-event curves for ISTH major bleeding and clinically relevant non-major bleeding of the Device and Control groups, respectively.

3.2.2 Sample Size for the Primary Safety Endpoint

This sample size was calculated employing log-rank test methodology using EAST 6 software with the following assumptions:

- Expected event rate of the Device group= 14% and of the Control group = 20%
- Alpha = 5% (two-sided)
- Power = 86.4%
- Expected attrition rate = 20%
- Required sample size = 1600 subjects

The expected primary safety endpoint event rate in the Device group is based off the combined PROTECT AF, CAP, PREVAIL, and CAP2 studies' WATCHMAN arms. The observed rate of non-procedural major bleeding at 36 months in these subjects was 9.4%. The primary safety endpoint event rate in the OPTION Device group is then expected to be 14%, which accounts for similar non-procedural major bleeding risk as in previous WATCHMAN studies but with the addition of clinically relevant non-major bleeds unrelated to the implant procedure. The expected rate of 20% in the Control group is based off the rates of major or clinically relevant non-major bleeds reported in the ARISTOTLE, ENGAGE, and ROCKET-AF studies.

Given the above assumptions, 1280 subjects will be required. In order to account for up to 20% expected rate of attrition which includes subjects who die, withdraw, or are lost-to-follow-up through 36 months without experiencing an endpoint event, a total of 1600 (1280/(1-0.2)) subjects need to be enrolled. A sample size of 1600 subjects provides approximately 86.4% power for the Primary Safety Endpoint analysis.

3.2.3 Statistical Methods for the Primary Safety Endpoint

All subjects in the randomized cohort will be included in the analysis. A log-rank test will be performed including all subjects in database at the time of analysis post 36 months. The objective is met if the p-value of the log-rank test is less than 0.05.

3.3 Secondary Endpoints

The secondary endpoint is ISTH major bleeding at 36 months (including procedural bleeding). This endpoint is defined as the Kaplan Meier estimates of time to first occurrence of major bleeding at 36 months and will be tested for non-inferiority of WATCHMAN to Control

3.3.1 Hypotheses

The objective of the secondary endpoint analysis is to test the null hypothesis that the difference in cumulative incidence between the Device and Control groups is greater than a prespecified noninferiority margin δ. The null hypothesis will be tested vs. the one-sided alternative hypothesis that the difference in cumulative incidence is less than a noninferiority margin δ.

H_0: $S_1(t) \geq S_0(t) + \delta$

H_a: $S_1(t) < S_0(t) + \delta$,

where $S_1(t)$ and $S_0(t)$ are the Kaplan Meier estimates for cumulative incidence of all ISTH major bleeding at 36 months for the Device and Control groups, respectively.

3.3.2 Sample Size for Secondary Endpoint

The power and sample size were calculated using Farrington Manning methods for non-inferiority using SAS version 9.4. The assumptions and parameters pertaining to this design are as follows:

- Power = 84.3%
- Alpha = 0.025 (one-sided)
- Expected Device and Control group performance: cumulative incidence of 11% at 36 months in both groups
- A noninferiority margin δ = 5.25%
- Attrition: cumulative attrition rate of 20% in the Device Group and in the Control Group at 36 months.

The expected device time-to-event rate of 11% at 36 months was derived from subjects with previous ablations from the combined PROTECT AF, CAP, PREVAIL, and CAP2 studies' WATCHMAN arms. The observed event rate in these subjects was 13.6% event at 3 years. This rate is expected to be lower in the OPTION device arm due to enhanced implant experience over time, device improvements, and the use of direct oral anticoagulants post-implant rather than warfarin. The expected event rate in the Control group is expected to be similar to that of the Device group. The non-inferiority margin of 5.25% represents a relative risk of 1.48.

Given the above assumptions, 1280 subjects will be required. In order to account for up to 20% expected rate of attrition which includes subjects who die, withdraw or are lost-to-follow-up through 36 months without experiencing an endpoint event, a total of 1600 (1280/(1-0.2)) subjects need to be enrolled.

A sample size of 1600 subjects provides approximately 84% power for the secondary endpoint analysis.

In addition, the power and sample sizes have been re-evaluated for non-inferiority testing settings that would match the method described to perform the hypothesis analysis, namely differences in the Kaplan-Meier event rates of the respective endpoints using the Greenwood formula.

The operating characteristics of the statistical test for the primary effective endpoint, and the secondary endpoint, are calculated by simulating 10,000 trials, for each endpoint, for a given sample size using custom-written code in SAS. The results are consistent with the sample size calculation performed using a parametric formula, namely the Farrington-Manning method above. The table below presents the simulation results for the secondary safety endpoint.

Method	**Farrington-Manning**	**Simulation**
Power	84.3%	88.9%
Alpha (Type I error)	0.025	0.0254

3.3.3 Statistical Methods for the Secondary Endpoint

All subjects in the randomized cohort will be included in the analysis. The secondary endpoint event from the WATCHMAN Device and Control arms, $S_1(t)$ and $S_0(t)$ at 36 months, defined as the 1095^{th} day post-randomization, will be estimated by the Kaplan Meier method. The 97.5% one-sided upper bound of confidence limit of the difference between WATCHMAN FLX and Control rates at 36 months will be calculated using the Greenwood formula for the variance of the Kaplan Meier estimates. The objective is met if this confidence limit is less than the predefined noninferiority margin of 5.25%.

3.3.4 Superiority Test for the Secondary Endpoint

If the non-inferiority hypothesis objective is met and the observed event rate favors the device arm, i.e., $S_1(t) < S_0(t)$, a superiority test will be subsequently performed and a one-sided p-value of 0.025 will be considered significant. No multiplicity adjustment is needed.[1]

3.3.5 Multiplicity Consideration

The strategy to avoid the multiplicity issue for the additional tests is to have a hierarchical order for the hypothesis tests and the testing significance level [1] for all hypothesis testing is one-sided alpha 0.025. The hypothesis testing will follow the order below.

1. Superiority testing of the primary safety endpoint and non-inferiority testing of the primary efficacy endpoint,
2. Non-inferiority testing of the secondary endpoint,
3. Superiority testing of the secondary endpoint and
4. Superiority testing of the primary efficacy endpoint.

The study is successful if Superiority testing of the primary safety endpoint and non-inferiority testing of the primary efficacy endpoint are met.

3.3.6 Statistical Methods

The primary efficacy endpoint and the secondary endpoint event rate, $P_1(t)$ and $P_0(t)$, respectively, will be estimated by the Kaplan Meier method. The timepoint of 36 months is defined as the 1095th day post-randomization. The 97.5% one-sided upper bound of confidence limit of the difference between Watchman FLX and Control rates will be calculated using Greenwood's formula for the variance of the Kaplan Meier estimates. The objective is met if this confidence limit is less than the predefined noninferiority margin of 5.0% for the primary effectiveness endpoint and 5.25% for the secondary endpoint.

In addition, calculation of an approximate p-value can be obtained. The critical value will be generated using the Com-Nougue approach to estimate the Z-statistic with Greenwood formula for estimating the variance.

The 97.5% one-sided confidence interval for event rate difference is

$$P_1(T) - P_0(T) \pm z_{1-\alpha/2}\sqrt{var(P_1(T)) + var(P_0(T))}$$

and

$$Z = \frac{P_1(T) - P_0(T) - \delta}{\sqrt{var(P_1(T)) + var(P_0(T))}}$$

P-value=1-probnorm(Z)

where for each treatment group (1=Device, 0=Control)

$P_1(T)$, $P_0(T)$ = Kaplan Meier estimator of event rate at time t=T=1095 day

δ = non-inferiority margin

Var= σ^2 =variance of the estimator where

σ = standard error by Greenwood's formula = $S(T)\sqrt{\sum_{j=1}^{T=1095} \frac{d_j}{n_j s_j}}$,

S(T)=KM freedom from event estimate at time t=T

$P_i(T)$=1-S(T), i=0,1

n = number of subjects

d = number of events

s=n — d for the j^{th} event time interval

3.4 Additional Endpoint Analyses

1. The occurrence and incidence of:
   - Stroke
     - Ischemic stroke
     - Hemorrhagic stroke

Disabling stroke/ Non-disabling stroke as defined using Munich Consensus[2]

- Systemic embolism
- Procedural and non-procedural bleeding and classifications; ISTH major bleeding and clinically relevant non-major bleeding
- All-cause death
    - Cardiovascular/unknown death
    - Non-cardiovascular death
- Device related thrombus

2. Device success (device successfully deployed and released) and procedural success (device successfully deployed and released, and absence of procedure related death)
3. Rates of effective (defined as jet size of ≤5mm) and complete (defined as no peri-device flow) LAA closure at 3- and 12-months post implant
4. Freedom from atrial fibrillation (Protocol Section 13 for definition)

Clinical recurrences of atrial fibrillation are defined as any of the following:

- Documented atrial fibrillation episode, or new onset of atrial flutter or atrial tachycardia event (≥ 30 seconds in duration or from a 10 second 12-lead EKG) post-randomization or index procedure, whichever is later and the end of study.
- Any of the following interventions for atrial fibrillation, or new onset of atrial flutter or atrial tachycardia post-randomization or index procedure, whichever is later and the end of study:
    - Repeat ablation procedure
    - Electrical and/or pharmacological cardioversion for atrial fibrillation/atrial flutter/atrial tachycardia
- Prescribed a higher dose of any anti-arrhythmic drug documented at baseline
- Prescribed a new anti-arrhythmic drug not documented at baseline
- Failure to discontinue AAD between 91 days and end of study post ablation procedure

5. Healthcare resource utilization
6. Quality of life

Device and procedural success will be assessed for Device subjects undergoing an intent to treat. The rates of effective and complete LAA closure at 3 and 12 months will be assessed for successfully implanted Device subjects with echo core lab measurements available from their respective follow-up TEE or equivalent. Device and procedural success and the rates of effective and complete LAA closure will both be summarized as a binary rate with exact 95% confidence intervals.

All time-to-event endpoints will be summarized using Kaplan-Meier methodology, and the time-to-event curves will be compared across treatment groups using a log-rank test. For the analysis of post-procedural major bleeding, the start time of follow-up in the Device group will be the day following the implant attempt, with the start time of follow-up in the Control being the day of randomization. In addition, the proportion of all strokes that are disabling will be compared across treatment groups using an exact binomial test.

All additional endpoint analyses will be exploratory in nature with no multiplicity adjustments, and will be analyzed once, at the time of primary efficacy and primary safety endpoint analyses.

4 GENERAL STATISTICAL METHODS

4.1 Analysis Sets

The primary analysis for the primary effectiveness endpoint will be done on an intent-to-treat basis, with each subject analyzed as being part of their randomized group regardless of the actual treatment received. The start time of follow-up for the intent-to-treat analysis will be the day of randomization. Additional analysis sets per actual treatment received will be performed for the primary and secondary endpoints as sensitivity analyses. It includes the following:

Treated as Assigned Analysis
This analysis set will include all randomized Device subjects who are successfully implanted as assigned. For these subjects, the start time of follow-up will be at the day of procedure. This cohort will also include all randomized Control subjects. The Control subjects who undergo any LAA closure procedure will be censored on the day of the LAA closure procedure.

The analysis of the primary safety endpoint will include all data from subjects randomized to the Device who undergo a **treated as assigned.** All secondary and additional endpoint analyses will be performed on an intent-to-treat basis.

4.2 Handling of Crossover

Control subjects will be allowed to "crossover" and receive a WATCHMAN device only after experiencing a primary efficacy or primary safety event endpoint event. These subjects will be censored from all analyses on the day after the event occurred. The primary efficacy and primary safety endpoints analysis will not be affected by this crossover as it uses time-to-event methodology, i.e. subjects no longer contribute to the analysis after experiencing an event. This is also true for any additional endpoint analysis that involves death, ischemic stroke or systemic embolism. Control subjects who undergo any LAA closure procedure prior to experiencing a primary efficacy or primary safety event endpoint event will be included in all analyses per the intent-to-treat principle.

4.3 Control of Systematic Error/Bias

Selection of subjects for enrollment will be made from the investigator's usual patient load. All subjects meeting the eligibility criteria and having signed the ICF will be eligible for enrollment in the study. To control for the potential bias that could be introduced via sponsor classification of endpoint events, an independent Clinical Events Committee (CEC) will adjudicate all potential endpoint events to be used in the data analyses.

4.4 Control of Type-I Error

The primary effectiveness endpoint will be tested at a significant level of 5% (two-sided, or equivalently 2.5% for one-sided test) using the methods described in Section 3.1, and the primary safety endpoint will be analyzed using a two-sided significance level of 5%. As the type-I error level is maintained below 5% for both the primary effectiveness and safety endpoints, the overall type-I error level across both the primary effectiveness and safety endpoints is maintained below 5%. This follows the methodology of the Intersection-Union Test (IUT).

4.5 Methods for Handling Missing Data

When calculating rates of adverse events for specific time intervals, missing and partial dates will be handled as shown below:

Partial Date Description	Action Taken
Entire onset date is missing	The implant procedure date will be used for the onset date for device patients and the randomization date will be used for Control patients.
The month and the day of the month are missing but the year is available	January 1 will be used for the month and day of the onset date. However, if the imputed date falls before the implant procedure date for Device patients and the Randomization date for Control patients , then the implant procedure date for Device patients and the Randomization date for Control patients will be used for the onset date.
Day is missing, but the month and year are available	The 1st will be used as the day of the onset date. However, if the imputed date falls before the implant procedure date for Device patients and the Randomization date for Control patients, then the implant procedure date for Device patients and the Randomization date for Control patients will be used for the onset date.

5 SUPPORTING DATA ANALYSES

The analyses described in this section will be performed at a time determined for the primary effectiveness and safety endpoint analysis. No adjustments for multiple comparisons will be made. Additional analyses will be performed as appropriate.

Data collected related to COVID-19 will be assessed. If there is significant impact of COVID-19 on results, additional analyses (e.g.: sensitivity) will be performed to assess the impact of COVID-19 on the study results according to regulatory guidance.

5.1 Description of Baseline Variables

Baseline demographic and clinical variables will be summarized for each of the treatment groups for both the intent-to-treat and per protocol populations. All continuous variables will be summarized as means, medians, standard deviations and interquartile ranges and compared between treatment groups using one-way analysis of variance (ANOVA) or the non-parametric Kruskal-Wallis test. Categorical variables will be summarized as frequencies and percentages and compared between treatment groups using Pearson's χ^2 test or Fisher's exact test, as appropriate. Ordinal variables will be compared using Cochran-Mantel-Haenszel test with row mean scores.

A by-subject listing of key demographic data and baseline characteristics, including variables used for stratification will be provided.

5.2 Subgroup Analyses

Primary effectiveness and safety endpoint results will be summarized. Additionally, treatment groups compared based on the following subgroups:

- Sequential vs Concomitant patients
- Male vs female
- Age at time of consent (< 75 years vs. ≥ 75 years)
- Stroke risk (CHA_2DS_2-VASc score 2-3, 4-5, and >5)
- Bleeding risk (HAS-BLED scores of 0, 1-2, and > 2)
- AF type at baseline (paroxysmal, persistent, permanent)
- Implanting image techniques (TEE vs CT for core lab data)

Time-to-event curves will be constructed for each treatment group, HRs and log-rank test p-values will be calculated within each subgroup, and subgroup by treatment interactions will be tested using Cox regression.

5.3 Univariate and multivariate Analyses

Univariate and multivariate Cox regression analyses will be used to assess the effect of baseline and post-procedural covariates on the primary effectiveness endpoint and primary safety endpoint. These analyses will be performed as ancillary to the primary endpoint analyses.

A univariate analysis will model the effect of each covariate within each treatment group. A multivariate analysis will model the time to ischemic stroke or SE on each covariate as well as treatment and covariate-by-treatment interaction. The covariates to be included in these modelling analyses are as follows:

- CHA_2DS_2-VASc score at baseline
- HAS-BLED score at baseline
- Concomitant vs sequential ablation
- Device size

For all of the modelling analyses mentioned in this section, continuous or ordinal measures may be dichotomized (e.g. age ≥75, LVEF <40, device size ≤24).

5.4 Pooling Analyses

The analyses of the primary effectiveness and safety endpoints will be presented using data pooled across investigational centers and geographic regions. Analyses will be performed to justify this pooling. Poolability will be assessed across site, geographical region, and sequential vs. concomitant ablation for each primary endpoint. Results will also be presented separately for site, geographical region, and sequential vs. concomitant ablations regardless of the results of the poolability assessments. Sites that successfully implanted with 5 subjects or fewer will be considered “small sites” and will be grouped together as one site per geographical region for the purpose of poolability assessment.

5.4.1 Pooling of Investigational Centers

Center-to-center heterogeneity will be assessed for the primary effectiveness and safety endpoint using a shared frailty model. Time-to-event endpoint will be modelled on randomized treatment group, with investigational center included as a random effect with a lognormal distribution. Centers will be deemed heterogeneous with respect to the primary effectiveness and safety endpoint if the p-value from Wald test for the investigational center random effect is <0.15.

5.5 Sensitivity Analyses for the Primary Endpoints

Per the Kaplan-Meier method, subjects who count towards attrition will still contribute towards the primary effectiveness endpoint event rates until the time at which they are lost to follow-up (i.e. censored). Sensitivity analyses will be performed as follows.

Regarding attrition bias, baseline characteristics will be compared across attrition vs non-attrition subjects for each primary endpoint to determine if subjects counted towards attrition had a higher baseline risk profile. In addition, tipping-point analysis will be conducted to determine the lowest/highest assumed rate in attrition subjects (assuming they had been followed for 3-years) such that the alternative hypothesis is no longer met/failed. The probability of observing this tipping-point rate in the attrition subjects (assuming a true rate equal to that observed in non-attrition subjects) will then be calculated to determine the likelihood of observing this tipping-point rate in the attrition subjects.

5.6 Missing Data

Every effort will be undertaken to minimize missing data. In time-to-event outcomes drop-outs will be censored at the time of last follow-up, consistent with the Kaplan-Meier methodology. The last follow-up date will be the latest of the following dates for each subject: date of adverse event, randomization date, implant procedure date, discharge date, and follow-up visit date. Unless otherwise specified, no statistical techniques will be used to impute missing data for continuous or categorical outcomes. The number of subjects included in each analysis will be reported so that the reviewer can assess the potential impact of missing data.

For each primary and secondary endpoint, a sensitivity analysis will be conducted to assess the impact of censored data and will include a worst-case analysis.

5.7 Quality of life

Clinical trials increasingly recognize the value of including patient-reported outcome measures in their design. To understand the impact of atrial fibrillation related procedures and disease management on patient's quality of life, the quality-of-life instruments used for the trial will be the EQ-5D-5L and the 12-Item Short Form Health Survey (SF-12) for generic instruments. The EQ-5D is a descriptive system of health-related quality of life states consisting of five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each of which can take one of five responses. The responses record five levels of severity (no problems/slight problems/moderate problems/severe problems/extreme problems) within a particular EQ-5D dimension. The SF-12 includes 12 dimensions with varying levels of response. Subjects will be asked to complete the questionnaires at the enrollment visit as well as at the 12- and 36-month follow-up visit.

5.8 Healthcare resource utilization

A formal health economics analysis may be completed as part of this trial, given meaningful clinical results are obtained. This will take into consideration any differences in survival, complication rates, quality of life, and resource utilization. The EQ-5D and SF-12 may be used to assess health utilities. We will capture health care utilization measures at all sites. These inputs may be used in health economics analysis performed beyond the scope of this Statistical Analysis Plan (SAP).

5.9 Interim Analyses

No formal interim analyses are planned for the purpose of stopping this study early for declaring effectiveness or for futility.

5.10 Changes to Planned Analyses

Any changes to the planned statistical analyses outlined in this plan will be documented in the clinical study reports along with a reason for any deviation.

6 VALIDATION

All clinical data reports generated per this plan will be validated per 90702587, Global WI: Clinical Data Reporting Validation.

7 PROGRAMMING CONSIDERATIONS

7.1 Statistical Software

Statistical analyses will be done using The SAS System Version 9.2 software or above (SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513, USA. All rights reserved.). Other statistical software (e.g., R) may be used as necessary.

7.2 Format of Output

Results of analysis will be output programmatically to Word documents from SAS with no manual intervention. All output for the final statistical report will be in the form of a Word document containing tables, figures, graphs, and listings, as appropriate.

7.3 Rules and Definitions

Binary event rates (proportions) will be calculated on a per patient basis. Two-sided exact 95% confidence intervals will be given for all applicable endpoints presented as proportions.

For all time-to-event analyses, event-free subjects will be censored at the date of their last follow-up. The last follow-up date will be the latest of the following dates for each subject: date of adverse event, randomization date, implant procedure date, discharge date, and follow-up visit date.

Days to (event or censoring date) = (event or censor) date – randomization date.

7.4 Clinical Events Committee

An independent Clinical Events Committee (CEC) will adjudicate all potential primary and secondary endpoints. CEC responsibilities, membership, meeting frequencies and procedures are outlined in the CEC charter. The CEC will make the final adjudication and classification of the

primary and secondary endpoints per the CEC charter, and the CEC determinations will supersede the site-reported data and sponsor classifications in all analyses of the primary and secondary endpoints.

7.5 Analyses of All Reported Serious Adverse Events

The proportion of subjects experiencing a serious adverse event (SAE), as well as the number of SAEs experienced, will be presented. These results will be presented both in aggregate as well as by SAE classification. The CEC classification will be used for events reviewed by the CEC, and the MEDRA LLT classification will be used for events not reviewed by the CEC.

8 REFERENCES

1. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604. doi:10.1002/sim.3495

2. Tzikas A, Holmes DR Jr, Gafoor S, et al. Percutaneous left atrial appendage occlusion: the Munich consensus document on definitions, endpoints, and data collection requirements for clinical studies. *EP Eur*. 2017;19(1):4-15. doi:10.1093/europace/euw141

9 APPENDICES

9.1 Expected Event

The Primary Effectiveness Endpoint analysis is a time-to-event analysis. The reference is provided below.

9.1.1 Historical event rates of WATCHMAN subjects with previous A-fib ablations

Event Rates for the primary effectiveness endpoint as well as the endpoint components are presented below. These are based on outcomes from patients with prior ablation in the PROTECT, PREVAIL, CAP, and CAP2 studies. Summary statistics for CHA_2DS_2-VASc and HAS-BLED scores are also provided as the baseline risk profile differed across historical studies.

Baseline Risk Scores and 3-Year Event Rates in Historical WATCHMAN Subjects with Previous AF Ablations

Measure	Overall (N=408)	PROTECT AF (N=70)	CAP (N=131)	PREVAIL (N=70)	CAP2 (N=137)
CHA_2DS_2-VASc Score	3.8±1.4 (407) (1.0, 8.0)	2.9±1.4 (69) (1.0, 8.0)	3.7±1.4 (131) (1.0, 7.0)	3.9±1.1 (70) (2.0, 7.0)	4.4±1.2 (137) (2.0, 8.0)
HAS-BLED Score	2.0±1.0 (408) (0.0, 5.0)	1.7±1.0 (70) (0.0, 4.0)	2.3±1.1 (131) (0.0, 5.0)	2.1±1.0 (70) (0.0, 4.0)	1.9±0.9 (137) (0.0, 5.0)
All-cause death, all stroke, or SE	**14.3% (53/370)**	8.5% (5/59)	13.0% (16/123)	17.2% (11/64)	16.9% (21/124)
All-cause death	9.5% (35/370)	5.1% (3/59)	10.6% (13/123)	6.3% (4/64)	12.1% (15/124)
Hemorrhagic stroke	0.3% (1/370)	0.0% (0/59)	0.8% (1/123)	0.0% (0/64)	0.0% (0/124)
Ischemic stroke	4.9% (18/370)	5.1% (3/59)	1.6% (2/123)	10.9% (7/64)	4.8% (6/124)
SE	0.3% (1/370)	0.0% (0/59)	0.0% (0/123)	0.0% (0/64)	0.8% (1/124)
Results are presented as mean +/- SD (N) (min, max) for continuous measures, or % (n/N) for binary rates. Event free subjects with <1035 days of follow-up are excluded from binary rate calculations.					

The event rate at 3 years as across all studies is 14.3%. For the OPTION study, we use the expected rate of 10%. The cumulative incidence is expected to be lower in the OPTION study due to enhanced implant experience over time, device improvements, and a decreased baseline risk profile for subjects enrolled in the OPTION study. The expected cumulative incidence in the Control arm is expected to be equal to that of the Device arm.

9.1.2 Expected cumulative incidence in the Control arm

Historical baseline risk factors and 3-year event rates are presented for both WATCHMAN and Control subjects below. The number of historical Control subjects with previous ablation is too small to draw conclusion about the expected performance of the OPTION Control arm.

Event Rates by Treatment Arm in Historical Subjects with Previous Ablation
All PROTECT AF, CAP, PREVAIL, and CAP2 Subjects with Previous Ablation

Measure	Historical WATCHMAN Study Subjects with Previous Ablations	
	WATCHMAN Subjects (N=408)	Control Subjects (N=68)
Median Age	72.5	71.6
Mean CHADS Score	2.3	2.3
Mean CHA2DS2VASc Score	3.8	3.5
Mean HASBLED Score	2.0	1.7
Paroxysmal AF	50.7% (207/408)	50.0% (34/68)
Persistent AF	29.4% (120/408)	26.5% (18/68)
Permanent AF	14.2% (58/408)	20.6% (14/68)
3-Year Event Rates		
All-cause death, all stroke, or SE	**14.3% (53/370)**	7.4% (4/54)
All-cause death	9.5% (35/368)	5.7% (3/53)
CV death, all stroke, or SE	8.6% (30/349)	3.8% (2/52)
CV death	3.5% (12/347)	2.0% (1/51)
Hemorrhagic stroke	0.3% (1/336)	2.0% (1/51)
Ischemic stroke	5.3% (18/339)	2.0% (1/51)
SE	0.3% (1/336)	0.0% (0/50)
Major bleeds (excluding bleeds <=45 days in WATCHMAN studies)	6.2% (21/339)	9.8% (5/51)
Event free subjects with <1035 days of follow-up are excluded from binary rate calculations.		

9.1.3 Expected primary safety event rate in the Device group

The rates of any non-procedure related major bleeding, as well as the rates of each non-procedure related major bleeding event type, are presented for historical WATCMAN subjects below. Clinically relevant non-major bleeding events are not included in this table as these events were not consistently collected in the historical WATCHMAN studies. The mean HAS-BLED scores for the WATCHMAN arm of each study are also provided as the baseline bleeding risk differed across studies.

36-Month Rates of Non-Procedure Related Major Bleeding in Historical WATCHMAN Subjects

Measure	Overall (N=1876)	PROTECT AF (N=463)	CAP (N=566)	PREVAIL (N=269)	CAP2 (N=578)
Mean HAS-BLED Score	2.04	1.78	2.30	2.04	2.00
Any Non-Procedural Major Bleed	**9.4% (160)**	**4.5% (18)**	**10.2% (53)**	**8.5% (21)**	**12.7% (68)**
Anemia Requiring Transfusion	0.7% (11)	0.0% (0)	0.4% (2)	0.9% (2)	1.3% (7)
Bruising - Hematoma	0.1% (1)	0.0% (0)	0.2% (1)	0.0% (0)	0.0% (0)
Cranial Bleed	0.5% (8)	0.5% (2)	0.0% (0)	0.4% (1)	1.0% (5)
Epistaxis	1.1% (19)	0.3% (1)	1.6% (8)	0.4% (1)	1.7% (9)
Gastrointestinal Bleeding	4.7% (80)	3.0% (12)	7.6% (40)	5.8% (14)	2.6% (14)
Hematoma	0.2% (3)	0.0% (0)	0.0% (0)	0.0% (0)	0.6% (3)
Hematuria	0.1% (1)	0.0% (0)	0.0% (0)	0.4% (1)	0.0% (0)
Major Bleed Requiring Transfusion	1.2% (21)	0.2% (1)	0.4% (2)	0.9% (2)	3.0% (16)
Pericardial Effusion with Cardiac Tamponade	0.1% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.2% (1)
Pseudoaneurysm	0.1% (1)	0.0% (0)	0.0% (0)	0.4% (1)	0.0% (0)
Stroke - Hemorrhagic	0.6% (9)	0.5% (2)	0.6% (3)	0.8% (2)	0.4% (2)
Subdural Hematoma	0.5% (8)	0.0% (0)	0.0% (0)	0.8% (2)	1.2% (6)
Results are presented as either mean HAS-BLED score, or KM event rate (# of subjects experiencing event)					

As shown above, the observed rate of non-procedural major bleeding events at 3 years with WATCHMAN was 9.4% in prior studies. Since the OPTION primary safety endpoint includes clinically relevant non-major bleeding events in addition to major bleeding events, the estimated primary event rate in the WATCHMAN arm has been increased to 14% at 36 months to account for the expected clinically relevant non-major bleeding events.

9.1.4 Operating characteristics of the proposed design

For the primary effectiveness endpoint, the sample size was determined using the Farrington-Manning method, while the primary analysis for this endpoint is based on the Kaplan-Meier estimates for the 36-months event rates and the Greenwood formula for the variance of the Kaplan-Meier estimates. A similar issue was noticed for the secondary endpoint. Additional simulations were conducted to ensure the operating characteristics of the proposed design, in particular the power and type I error rate for the non-inferiority testing of the primary effectiveness endpoint as well as for the testing of the secondary endpoint. Operating characteristics of the proposed design, in particular the power and type I error rate for the non-inferiority testing of the primary effectiveness endpoint as well as for the testing of the secondary endpoint are evaluated as follows.

Per communication with FDA, the power and sample sizes for the primary effectiveness and secondary endpoints have been re-evaluated. BSC recalculated the power and type I error rate for the non-inferiority testing settings in order to match the methods described to perform the

hypothesis analysis, specifically, differences in the Kaplan-Meier event rates of the respective endpoints using the Greenwood formula.

The operating characteristics of the statistical test for the primary effective endpoint, and the secondary safety endpoint, are calculated by simulating 10,000 trials, for each endpoint, for a given sample size using custom-written code in SAS. These SAS codes are provided along with this response to FDA.

The results are consistent with the sample size calculation performed using a parametric formula, namely the Farrington-Manning method. In particular,

Primary Effective Endpoint

	Protocol 1 (Before)	**Protocol Rev B/C (After)**
Method	**Farrington-Manning**	**Simulation, as FDA advised**
Power	85.9%	89.3%
Alpha (Type I error)	0.025	0.024

Secondary Safety Endpoint

	Protocol 1 (Before)	**Protocol Rev B/C (After)**
Method	**Farrington-Manning**	**Simulation, as FDA advised**
Power	84.3%	88.9%
Alpha (Type I error)	0.025	0.0254

Retaining the Type I error at 0.025 and its proximity, the power for the primary endpoint and the secondary endpoint under the simulation method exceed those of the Farrington-Manning method. Therefore, the original calculations hold as being more conservative. For the secondary safety endpoint, the difference in simulated type I error vs alpha is 0.0004, which is 1.6% of the nominal value of Alpha=0.025. As explainable in the statistical principles, the slight gain in alpha is offset by a corresponding gain in power (1- Beta), in this case 5.6% increase in the power estimate by the simulation method; therefore, the control of overall type I and type II errors remain equivalent and have not degraded. The SAS codes are in the attachments.

The original sample size method and statement remain valid.

10 ATTACHMENTS: SAS SIMULATION CODES

```
/***
Primary Effectiveness NI Power Simulations.sas

    Simulate power [and type I error] for a 2-arm study testing NI of treatment vs control
    Uses the KM event rates and Greenwood SE estimates in the Wald formula for testing NI in
       binary event rates.
***/

*** Power Simulation Study
The goal of this study is to estimate power when the true rate for the treatment group is equal
to the control, i.e. pT=pC. Roebruck and Kühn (1995) refer to this scenario as maximal power;

option compress=yes;
proc datasets library=work kill; run; quit;

%let ni = 0.05;                          * NI margin for difference in event rates;
%let nsim = 10000;             * number of simulated studies;

data sim_SE1;
      seed =20121217;  * Seed;*Seed1 and Seed2 necessary for grp1;
      seed2=20121218;  * Seed;*Seed1 and Seed2 necessary for grp2;
      t = 3;                     * Timepoint;

      N = 1600;                          * Overall sample size;
      attrition = 0.15;             * Attrition rate at time point of interest;

          F_t = 0.10;           * Assumed KM event rate in both arms [this may be random], Rx arm;
      S_t = 1 - F_t;   * Survival rate at time point of interest;
      lambda = -log(S_t)/t;
      lambda_attrition = -log(1-attrition)/t;

          F_t2 = 0.10;          * Assumed KM event rate in both arms [this may be random], control
arm;
      S_t2 = 1 - F_t2; * Survival rate at time point of interest;
      lambda2 = -log(S_t2)/t;
      lambda_attrition2 = -log(1-attrition)/t;

      do study = 1 to &nsim;
            do pt = 1 to N;
                              if mod(pt,2)=0 then grp=2; else grp=1;
                              if grp=1 then do;
                  raw_eventt = ranexp(seed)/lambda;
                  raw_attritiont = ranexp(seed)/lambda_attrition;
                              end;
                              if grp=2 then do;
                  raw_eventt = ranexp(seed2)/lambda2;
                  raw_attritiont = ranexp(seed2)/lambda_attrition2;
                              end;

                  if raw_eventt le raw_attritiont then do;
                        raw_time = raw_eventt;
                        raw_event = 1;
                  end;
                  else do;
                        raw_time = raw_attritiont;
                        raw_event = 0;
                  end;
                  if raw_time gt t then do;
                        event = 0;
                        time = t + 0.001;
                  end;
                  else do;
                        event = raw_event;
                        time = raw_time;
                  end;
                              if time<=t and event=0 then censor=1; else censor=0;
            output;
            end;
```

```
      end;
run;


option nonotes;
ods exclude all;
proc lifetest data=sim_SE1 outsurv=temp1_surv stderr  ;
      by study;
         strata grp;
      time time*event(0);
run;
ods exclude none;
option notes;

data temp2_surv;
      set temp1_surv;
      retain F SE;
      if survival ne . then F = 1 - survival;
      if SDF_STDERR ne . then SE = SDF_STDERR;
run;

data surv_SE1;
      set temp2_surv;
      by study grp;
      if last.grp;
run;

proc transpose data=surv_se1 out=surv_se_t1 prefix=f;
  by study;
  var f;
  id grp;
proc transpose data=surv_se1 out=surv_se_t2 prefix=se;
  by study;
  var se;
  id grp;
run;

data sim_final;
  merge surv_se_t1 surv_se_t2;
  by study;
  * calculate the upper 97.5% CI and determine if it's less than the NI margin;
  uci = (f1-f2) + 1.96*sqrt(se1**2+se2**2);
  pass = (uci < &NI); *PASS := Power ;
run;

proc means data=sim_final n mean maxdec=3;
      var pass ;
         *output out=sim_finalo;
      title1 'Endpoint Power';
         title2 'N=number of simulations, MEAN=Power';
run;
```

```
/***
Primary Effectiveness NI Type I error Simulations.sas
Simulate [power and] type I error for a 2-arm study testing NI of treatment vs control
        Uses the KM event rates and Greenwood SE estimates in the Wald formula for testing NI in
binary event rates. ***/

***Type I Error Simulation Study
Type I error is assessed by setting the true rate for the treatment group to be equal to the
respective noninferiority margin, i.e. on the boundary of the null hypothesis parameter space,
where the type I error rate is maximized. This means the probability of success for the treatment
group was calculated as follows: pT=pC+ δ (δ defined as noninferiority margin);

option compress=yes;
proc datasets library=work kill; run; quit;

%let ni = 0.05;                         * NI margin for difference in event rates;
%let nsim = 10000;                      * number of simulated studies;


data sim_SE1;
      seed =20121217;  * Seed;*Seed1 and Seed2 necessary for grp1;
      seed2=20121218;  * Seed;*Seed1 and Seed2 necessary for grp2;
      t = 3;                    * Timepoint;

      N = 1600;                         * Overall sample size;
      attrition = 0.15;           * Attrition rate at time point of interest;

         F_t = 0.15;          * Assumed KM event rate in both arms [this may be random], Rx arm;
      S_t = 1 - F_t;   * Survival rate at time point of interest;
      lambda = -log(S_t)/t;
      lambda_attrition = -log(1-attrition)/t;

         F_t2 = 0.10;         * Assumed KM event rate in both arms [this may be random], control
arm;
      S_t2 = 1 - F_t2; * Survival rate at time point of interest;
      lambda2 = -log(S_t2)/t;
      lambda_attrition2 = -log(1-attrition)/t;

      do study = 1 to &nsim;
            do pt = 1 to N;
                              if mod(pt,2)=0 then grp=2; else grp=1;
                              if grp=1 then do;
                  raw_eventt = ranexp(seed)/lambda;
                  raw_attritiont = ranexp(seed)/lambda_attrition;
                              end;
                              if grp=2 then do;
                  raw_eventt = ranexp(seed2)/lambda2;
                  raw_attritiont = ranexp(seed2)/lambda_attrition2;
                              end;

                  if raw_eventt le raw_attritiont then do;
                        raw_time = raw_eventt;
                        raw_event = 1;
                  end;
                  else do;
                        raw_time = raw_attritiont;
                        raw_event = 0;
                  end;
                  if raw_time gt t then do;
                        event = 0;
                        time = t + 0.001;
                  end;
                  else do;
                        event = raw_event;
                        time = raw_time;
                  end;
                              if time<=t and event=0 then censor=1; else censor=0;
            output;
            end;
```

```
      end;
run;


option nonotes;
ods exclude all;
proc lifetest data=sim_SE1 outsurv=temp1_surv stderr ;
      by study;
         strata grp;
      time time*event(0);
run;
ods exclude none;
option notes;

data temp2_surv;
      set temp1_surv;
      retain F SE;
      if survival ne . then F = 1 - survival;
      if SDF_STDERR ne . then SE = SDF_STDERR;
run;

data surv_SE1;
      set temp2_surv;
      by study grp;
      if last.grp;
run;

proc transpose data=surv_se1 out=surv_se_t1 prefix=f;
  by study;
  var f;
  id grp;
proc transpose data=surv_se1 out=surv_se_t2 prefix=se;
  by study;
  var se;
  id grp;
run;

data sim_final;
  merge surv_se_t1 surv_se_t2;
  by study;
  /*** NEED TO CHECK THAT THIS FORMULA IS CORRECT ***/
  * calculate the upper 97.5% CI and determine if it's less than the NI margin;
  uci = (f1-f2) + 1.96*sqrt(se1**2+se2**2);
  pass = (uci < &NI); *PASS := Type I error ;
run;

proc means data=sim_final n mean maxdec=3;
      var pass ;
         *output out=sim_finalo;
      title1 'Endpoint Type I error';
         title2 'N=number of simulations, MEAN=Type I error';
run;
```

Secondary Endpoint NI Power Simulations.sas

```
/*** Simulate power [and type I error] for a 2-arm study testing NI of treatment vs control
        Uses the KM event rates and Greenwood SE estimates in the Wald formula for testing
        NI in binary event rates. ***/

*** Power Simulation Study
The goal of this study is to estimate power when the true rate for the treatment
group is equal to the control, i.e. Pt = Pc. Roebruck and Kühn (1995) refer to this scenario as
maximal power. ;
option compress=yes;
proc datasets library=work kill; run; quit;

%let ni = 0.0525;             * NI margin for difference in event rates;
%let nsim = 10000;            * number of simulated studies;

data sim_SE1;
      seed =20121217;  * Seed;*Seed1 and Seed2 necessary for grp1;
      seed2=20121218;  * Seed;*Seed1 and Seed2 necessary for grp2;
      t = 3;                  * Timepoint;

      N = 1600;                       * Overall sample size;
      attrition = 0.20; * Attrition rate at time point of interest;

          F_t = 0.11;         * Assumed KM event rate in both arms [this may be random], Rx arm;
      S_t = 1 - F_t;   * Survival rate at time point of interest;
      lambda = -log(S_t)/t;
      lambda_attrition = -log(1-attrition)/t;

          F_t2 = 0.11;        * Assumed KM event rate in both arms [this may be random], control
arm;
      S_t2 = 1 - F_t2; * Survival rate at time point of interest;
      lambda2 = -log(S_t2)/t;
      lambda_attrition2 = -log(1-attrition)/t;

      do study = 1 to &nsim;
            do pt = 1 to N;
                              if mod(pt,2)=0 then grp=2; else grp=1;
                              if grp=1 then do;
                raw_eventt = ranexp(seed)/lambda;
                raw_attritiont = ranexp(seed)/lambda_attrition;
                              end;
                              if grp=2 then do;
                raw_eventt = ranexp(seed2)/lambda2;
                raw_attritiont = ranexp(seed2)/lambda_attrition2;
                              end;

                if raw_eventt le raw_attritiont then do;
                      raw_time = raw_eventt;
                      raw_event = 1;
                end;
                else do;
                      raw_time = raw_attritiont;
                      raw_event = 0;
                end;
                if raw_time gt t then do;
                      event = 0;
                      time = t + 0.001;
                end;
                else do;
                      event = raw_event;
                      time = raw_time;
                end;
                              if time<=t and event=0 then censor=1; else censor=0;
            output;
            end;
      end;
run;
```

```
option nonotes;
ods exclude all;
proc lifetest data=sim_SE1 outsurv=temp1_surv stderr  ;
      by study;
          strata grp;
      time time*event(0);
run;
ods exclude none;
option notes;

data temp2_surv;
      set temp1_surv;
      retain F SE;
      if survival ne . then F = 1 - survival;
      if SDF_STDERR ne . then SE = SDF_STDERR;
run;

data surv_SE1;
      set temp2_surv;
      by study grp;
      if last.grp;
run;

proc transpose data=surv_se1 out=surv_se_t1 prefix=f;
  by study;
  var f;
  id grp;
proc transpose data=surv_se1 out=surv_se_t2 prefix=se;
  by study;
  var se;
  id grp;
run;

data sim_final;
  merge surv_se_t1 surv_se_t2;
  by study;

  * calculate the upper 97.5% CI and determine if it's less than the NI margin;
  uci = (f1-f2) + 1.96*sqrt(se1**2+se2**2);
  pass = (uci < &NI); *PASS := Power ;
run;

proc means data=sim_final n mean maxdec=3;
      var pass ;
          *output out=sim_finalo;
      title1 'Endpoint Power';
          title2 'N=number of simulations, MEAN=Power';
run;
```

Secondary Endpoint NI Type I error Simulations.sas

```
/*** Simulate [power and] type I error for a 2-arm study testing NI of treatment vs control
        Uses the KM event rates and Greenwood SE estimates in the Wald formula for testing
        NI in binary event rates. ***/

***Type I Error Simulation Study
Type I error is assessed by setting the true rate for the treatment group to be equal
to the respective noninferiority margin, i.e. on the boundary of the null hypothesis parameter
space, where the type I error rate is maximized. This means the probability of success for the
treatment group was calculated as follows: Pt = Pc+ del (del  defined as noninferiority margin
in ). ;
option compress=yes;
proc datasets library=work kill; run; quit;

%let ni = 0.0525;             * NI margin for difference in event rates;
%let nsim = 10000;            * number of simulated studies;

data sim_SE1;
      seed =20121217;  * Seed;*Seed1 and Seed2 necessary for grp1;
      seed2=20121218;  * Seed;*Seed1 and Seed2 necessary for grp2;
      t = 3;                  * Timepoint;

      N = 1600;                       * Overall sample size;
      attrition = 0.20; * Attrition rate at time point of interest;

         F_t = 0.11 + &ni;            * Assumed KM event rate in both arms [this may be random],
Rx arm;
      S_t = 1 - F_t;   * Survival rate at time point of interest;
      lambda = -log(S_t)/t;
      lambda_attrition = -log(1-attrition)/t;

         F_t2 = 0.11;         * Assumed KM event rate in both arms [this may be random], control
arm;
      S_t2 = 1 - F_t2; * Survival rate at time point of interest;
      lambda2 = -log(S_t2)/t;
      lambda_attrition2 = -log(1-attrition)/t;

      do study = 1 to &nsim;
            do pt = 1 to N;
                              if mod(pt,2)=0 then grp=2; else grp=1;
                              if grp=1 then do;
                  raw_eventt = ranexp(seed)/lambda;
                  raw_attritiont = ranexp(seed)/lambda_attrition;
                              end;
                              if grp=2 then do;
                  raw_eventt = ranexp(seed2)/lambda2;
                  raw_attritiont = ranexp(seed2)/lambda_attrition2;
                              end;

                  if raw_eventt le raw_attritiont then do;
                        raw_time = raw_eventt;
                        raw_event = 1;
                  end;
                  else do;
                        raw_time = raw_attritiont;
                        raw_event = 0;
                  end;
                  if raw_time gt t then do;
                        event = 0;
                        time = t + 0.001;
                  end;
                  else do;
                        event = raw_event;
                        time = raw_time;
                  end;
                              if time<=t and event=0 then censor=1; else censor=0;
            output;
            end;
```

```
      end;
run;


option nonotes;
ods exclude all;
proc lifetest data=sim_SE1 outsurv=temp1_surv stderr ;
      by study;
         strata grp;
      time time*event(0);
run;
ods exclude none;
option notes;

data temp2_surv;
      set temp1_surv;
      retain F SE;
      if survival ne . then F = 1 - survival;
      if SDF_STDERR ne . then SE = SDF_STDERR;
run;

data surv_SE1;
      set temp2_surv;
      by study grp;
      if last.grp;
run;

proc transpose data=surv_se1 out=surv_se_t1 prefix=f;
  by study;
  var f;
  id grp;
proc transpose data=surv_se1 out=surv_se_t2 prefix=se;
  by study;
  var se;
  id grp;
run;

data sim_final;
  merge surv_se_t1 surv_se_t2;
  by study;

  * calculate the upper 97.5% CI and determine if it's less than the NI margin;
  uci = (f1-f2) + 1.96*sqrt(se1**2+se2**2);
  pass = (uci < &NI); *PASS := Type I error ;
run;

proc means data=sim_final n mean maxdec=4;
      var pass ;
         *output out=sim_finalo;
      title1 'Endpoint Type I error';
         title2 'N=number of simulations, MEAN=Type I error';
run;
```